Doktori (PhD) értekezés

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

TARTALOMJEGYZÉK

BEVEZETŐ

Bár főleg gépekkel kapcsolatos ez a kutatás, de kerülöm, hogy ezt gépiesen tegyem, és valami távoli, tőlem független objektumként kezeljem a feldolgozott anyagot. Igyekszem a tudományosság kritériumai mellett úgy vizsgálódni és a meglátásaimat úgy közreadni, hogy azáltal a tisztelt olvasó az új információk és gondolatok mellett a világunkhoz, saját megismeréséhez, önnön emberi lényegéhez is közelebb juthasson.

A KUTATÁS BEMUTATÁSA

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

A cím kifejtésével jól összefoglalható jelen kutatás programja, érdemes felvezetésként ezt ismertetnem. A mesterséges intelligencia (továbbiakban MI) kifejezés alatt első megközelítésben elegendő azt érteni, amit a tisztelt olvasó tud róla – tisztázása úgyis a célok között szerepel. A „humán aspektusok” kifejezés arra utal, hogy a téma a filozófiai etikától a pszichológiáig számos humántudománnyal kapcsolatba kerül, ez adja a kutatás interdiszciplináris jellegét. Tisztázandó, hogy katonaként miért nem a „honvédelem” kifejezést használom a címben, miért utalok egy tágabb területre? A válasz, hogy a katonai kifejezések (hadi, katonai, harci stb.) használata nem jól fedné le a kutatás szemléletét, sőt félrevezető lenne, amint erre majd maga a kutatás is rámutat. Olyan szót kerestem, amely a hagyományos katonai szegmensen túlmutatva utal az ország védelmének résztvevőire. Arra jutottam, hogy a „védelem” kifejezés az erre megfelelő kompromisszum, és annak ellenére is jó választás, hogy a köznyelvi „védelmi szféra” bizonyos részeit a tanulmány nem kívánja vizsgálni (ld. a téma lehatárolásánál).

Összefoglalva: az MI, a humán aspektus és a védelmi szféra kifejezések közös halmaza ugyan túlmutat az értekezés keretein, metszetük azonban jól utal a kutatás tartalmára. Még így is a címben megjelölt területnek is csupán néhány kis szeletét elemezhetem jelen keretek között.

A kutatásokat 2024. december elsején lezártam. Ezután célzott kutatásokat már nem folytattam, de az anyag kidolgozása közben néhány frissebb információt fontosságuk miatt kénytelen voltam a szövegbe dolgozni. Naprakész szöveg nem volt (és nem is lehetett volna) a célom egy ilyen gyorsan változó területen, de törekedtem minél időtállóbb megfogalmazásokra.

A téma fontossága, időszerűsége, valamint motivációja

Az MI vizsgálatának jelentősége ugyan vitathatatlan, viszont tisztázandó, hogy miben aktuálisak és lényegesek-e ennek humán aspektusai, illetve az ilyen megközelítés miért lényeges a

védelmi szférában. Azzal kezdem, hogy a téma nem pusztán időszerű, hanem kifejezetten „most van itt az ideje”. Egyrészt azért, mert az MI terjedésével már ki tud rajzolódni a társadalommal alakuló kapcsolata, tehát napjainkban a maga dinamikájában vizsgálható az, ami tíz éve még nem volt az. Másrészt a korábbi technológiák társadalmi (kölcsön-)hatásának fényében vizsgálódhatok, ami nagy előny, hiszen jelen technológia kifejlődésének olyan korai fázisában fogalmazhatok meg ajánlásokat, amire a korábbi technikai forradalmak esetében nem volt lehetőség. A kihívások ugyan radikálisan mások, viszont megfelelő szintű elvonatkoztatással pl. a motorizáció vagy a telekommunikáció társadalmi, környezeti, etikai, fejlődési stb. mintázatai jól adaptálhatóak a jelenre, sőt a virtuális és a kognitív térben zajló változásokra is. Vagyis elvileg jobb esélyünk van elébe menni az MI negatív hatásainak, és jól kihasználni a benne rejlő potenciálokat, mint a korábbi diszrupciók esetében volt. A megértés mellett egy ilyen kutatás segíthet tervezhetőbbé is tenni a védelmi szférában azt a paradigmaváltást, mely úgyis elkerülhetetlen, hiszen a világ minden részét érinti.

Az időszerűséggel kapcsolatban leírtak a téma fontosságának indoklásaként is értelmezhetőek, de nézzük meg a cím részeit itt is külön-külön. A humán aspektusok jelentőségét megvilágítja, ha arra gondol az olvasó, hogy az MI olyan fogalmakkal függ össze, mint az okosság, a tanulás, az érzelmek vagy a szabadság (autonómia). Ezek gépi utánzása annyira új, hogy egy sor, eddig ismeretlen problémát vethet fel, melyeket szükséges a tudományos-fantasztikus szerzőknél tudományosabban is vizsgálni. A védelmi aspektus jelentőségét sem szükséges túlmagyarázni, hiszen kézenfekvő, hogy minden jó találmánynak van árnyoldala, vissza lehet vele élni, tehát kockázatot jelent. Az MI esetében ezt erősíti, hogy egyre több állami feladat digitalizálódik. Az ilyen vizsgálatok azért is fontosak, hogy félelmek helyett a lehetőségek helyes használatát segítsék előtérbe kerülni.

Bemutatom személyes belső motivációmat is, mely segítheti az olvasót a megközelítéseim hátterének megismerésében. Számomra a fő eredmény, hogy ebben a munkában sikerült végre egyesítenem érdeklődési köreim három fontos irányának korábbi széttagoltságát. Az első a tudományos vívmányok, a technológiák, elsősorban az informatikai megoldások iránti érdeklődésem (aminek mentén első egyetemi diplomámat szereztem informatika szakon). A második az a belső igény, amely 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, s amely idővel katonai hivatássá érett bennem. 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 jelen szintézise az olvasók számára is inspiratív és érdekes lesz.

A kutatás alapproblémája

A kutatás egészének általános problémakörét így lehet röviden megfogalmazni:

> *Az emberek közötti és az emberi közösségek (országok) közötti viszonyra nagy hatással vannak a technológia nagy lépései, ám az MI várható hatása ezek közül is kiemelkedik kognitív sajátosságainak újszerűsége miatt. Ennek konkrét vizsgálatához szükséges az országok közötti viszonyra ható MI tanulmányozása (ez a probléma védelmi aspektusa), illetve az emberek közötti információátadásra ható MI elemzése (ez adja a probléma humán és terminológiai tényezőit).*

Tehát jelen munka az MI védelmi célú további hatásvizsgálatai közé illeszkedik, és eredményeivel az ilyen irányú vizsgálatokat is támogatni szeretné. Bár a fenti probléma-megjelölés már előre vetíti a kutatás főbb irányait, a kijelölt terület még túlságosan széles. Ezt csak részben tudom szűkíteni majd a kutatási célok és kutatási kérdések által, mivel azok nem hivatottak tartalmazni, hogy milyen szempontok alapján vizsgálódik a szerző, vagy, hogy a közreadott megközelítések mely irányokból kívánják feldolgozni a problémákat (és milyen irányokból nem). Az utóbbi jellemzők rögzítése a kutatás tervezését is segíti, erre térek itt rá.

A vizsgálódás szempontjai

A kutatásban az alábbi szempontokat igyekeztem kiemelten szem előtt tartani.

1. **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 össze más tudományterületekkel, szorosabban a hadtudományok, a szociológia, illetve a tudománytörténet egyes részeivel.
2. A **védelmi aspektusok** érintése minél több részkutatásban.
3. T**erminológiai tisztázás.** Minden kutatási területen a fontos vagy kevésbé ismert kulcsfogalmak alapos magyarázata (különféle értelmezési módjaikkal együtt), a felmerülő terminológiai anomáliák tisztázása.
4. **Hiánypótlás**. Minél több olyan részterület bemutatása, amelyről magyarul nem jelentek meg publikációk, vagy esetleg világviszonylatban is kevés a forrás.

5. **Gyakorlatiasság a részletekben**. Nem csupán a tervezett tudományos eredmények, hanem a kisebb részeredmények magukban is legyenek hasznosak (pl. saját fogalmak, modellek, tagolások, egyedi megközelítések, meglátások, ötletek, javaslatok stb.).
6. **Előtérben a háttér**. A vizsgálat ne csak az MI által generált jelenségekre összpontosítson, hanem ezek lehetséges hátterére is.
7. **Prevenció**. A téma időszerűségénél fentebb leírtak alapján ez is kiemelt szempont.
8. **Egzisztenciális megközelítés**. Próbálok úgy közelíteni a témákhoz, hogy a vizsgálatok minél több eredményét alkalmazhassa az olvasó a saját életében is.

A téma lehatárolása (amit a kutatás kerül)

Ezt az interdiszciplináris témát multidiszciplinárisan lenne igazán hatékony megközelíteni, ám az emberi képességek határai miatt egy valódi össztudományi tudás nem megvalósítható. A felvezetésben megadtam azokat a tudományokat, melyekre leginkább támaszkodni kívánok, alább kizárom, hogy a téma kézenfekvő kapcsolódásai közül mely tudományok és kutatási területek azok, amelyeket csak annyiban érintek, amennyiben szükséges.

1. Jogi aspektusok – kiemelendő, hogy az etikát én filozófiai szempontból közelítem, nem a gyakoribb, jogtudományi értelemben használom a kifejezést. Bár szabályzási kérdésekkel foglalkozom, de ennek csupán információbiztonsági és védelmi aspektusait vizsgálom.
2. A védelem katasztrófavédelmi, rendészeti, titkosszolgálati, vízügyi stb. oldalainak vizsgálatára a felvezetésben leírtak szerint nem koncentrálok.
3. A gazdaságtudományt némely vizsgált rész érinti, de azt sem közgazdászként elemzem.
4. Az orvosi és biológiai tudományok felől csak a teljesség kedvéért, alapszinten közelítek a témához, az agykutatás, a biológiai rendszerek, vagy az emberbe épített MI-implantátumok alapos elemzése nem célom.
5. A kognitív hadszíntér – későbbi alapos vizsgálatát tervezem, itt csupán érintem.

PROBLÉMAKÖRÖK, CÉLOK ÉS HIPOTÉZISEK

Fentebb, a kutatás alapproblémájának ismertetésében előrevetítettem a kutatás fő irányait. Alább a két fő problémakör (P1-P2) szerint tagolva olvashatóak a kutatási célok (C1-C2) és a hipotézisek (H1-H2), melyek kitűzésénél a várható felhasználási lehetőségeket is figyelembe vettem, és ezeket fel itt is tüntetem. A 2. és 3. számú kutatási szempont alapján a terminológiai és védelmi vizsgálati attitűd a konkrét célokon túl általánosságban is áthatja a dolgozatot.

P1: A téma terminológiai problémakörei

C1: Minél több tényezőt azonosítani, melyek nélkül az MI meghatározása félreérthető, vagy amelyeket beleértve a fogalom zavarossá válik, valamint javaslatokat tenni a feltárt problémák feloldására (új fogalmak, felosztások, más kifejezések stb.).

H1: *A Mesterséges Intelligencia jelenlegi meghatározásaiból fontos aspektusok hiányoznak, és félreértések terhelik.*

Tervezett felhasználások: A kimutatott hiányosságok alapján az MI új, használhatóbb meghatározása, mely talán a következő tíz évben is megállja a helyét. Ezáltal kezelhetőbbé válnak a szabályozások, valamint a védelmi tervezés számára az MI olyan aspektusai, melyeket a jelenlegi meghatározások nem vesznek figyelembe. A jobb definíció segíti az MI lényegének megismertetését és a használatához szükséges szemlélet kialakítását is, mely az oktatás számára is kulcsfontosságú, az ilyen képzés pedig stratégiai jelentőségű.

P2: A védelmi paradigmaváltás problematikája

C2: A világ változásának, valamint az erőérvényesítés új tendenciáinak MI-vel való kapcsolatát elemezve határozni meg az MI védelmi szempontból fontos tényezőit.

H2: *Az MI tovább fokozza és támogatja azt a tendenciát, melynek során az erőérvényesítésben folytatódik hangsúlyeltolódás a puha műveletek felé*.

Tervezett felhasználások: Elsősorban a diszciplináris és stratégiai tervezés során lehet majd hasznosítható az eredmény, de a hozzá vezető út során kidolgozott terminológiák, modellek és meglátások hadtudományi szempontból is szükségesek.

A KUTATÁS MEGVALÓSULÁSI TERVE

A fenti célok alapján határoztam meg a vizsgálatok konkrét kérdéseit. Ezekhez igazodtak a kutatási módszerek, valamint a felhasznált szakirodalmak köre.

A K1-K6 kutatási kérdés és az értekezés logikai felépítése

A kutatás tervezésekor világossá vált, hogy a tervezett eredmények több irányú alátámasztására lesz szükség. Ennek logikus elrendezése külön kihívás volt, hiszen a vizsgálatok szorosan összekapcsolódnak, mindegyik függ a többitől. (Hogy segítsem az olvasót a sok belső összefüggés áttekintésében, az utalásokat pontos alfejezetszámokkal tettem követhetőbbé.) Végül a kifejtés elrendezésének sokféle logikus módja közül az alábbi tűnt a legkevesebb előreutalást igénylő verziónak, a fejezetarányossági elvárást is betartva. Az anyagot úgy tagoltam, hogy az első négy fejezet mindkét célra irányul, végül az V. fejezet a P1 problémakört pótolja ki, összesíti és rendszerezi, majd a VI. fejezetben P2 tekintetében érnek össze a gondolati szálak.

A vizsgálatok legelején be kellett mutatni, mi is az a dolog, amit vizsgálok, ezért az I. fejezet áttekinti az intelligencia kérdéskörét. A II. fejezetben a gépi tanulást, mint az intelligencia alapját vizsgálom meg. Ez kicsit technikai jellegű fejezet, ide került az MI-hez kapcsolódó technikák vázolása is, de a célközönség alapján igyekeztem kerülni a túlságosan műszaki mélységeket. A tanulás és az intelligencia következménye a szabadság valamilyen szintje. A szabadság, az egyéni és a közösségi autonómiák sok szállal kapcsolódnak a védelmi kihívásokhoz, ehhez társul korunkban a gépek esetleges autonómiája. Ezért a III. fejezetben az autonómia kérdéskörét és a vele kapcsolatos gondolataimat foglaltam össze. Az MI-hez kapcsolódó három alapfogalom vizsgálata után, – ezekre építve – a IV. fejezet kritikusan tekinti át azokat a biztonsági kihívásokat, melyek a neurális hálóból fakadnak. De itt kapott helyet az MI fejlődési és terjedési lehetőségeinek vizsgálata is. Az V. fejezetben a C1 cél tervezett elérése mellett tovább is kívánok lépni: még a terminológiai kérdéskör részeként volt logikus ugyanis tárgyalni az informatika kifejezés várható változását az MI fényében, melyet az utolsó vizsgálat során használok fel. A VI. fejezet részkutatásai, a teljes addigi anyagra épülve kifejezetten védelmi témákat tárgyalnak a C2 cél elérése érdekében. A bemutatott felépítés alapján határoztam meg a kutatási kérdéseket, melyek a fejezetek konkrét programját jelölik ki:

K1: Az intelligencia kérdésköre: *Hogyan értelmezik a világban a mesterséges és az emberi intelligenciát, miként fogalmazzák meg ezeket, és hiányoznak-e ebből fontos aspektusok?*

K2: A technológia és tanulás kérdésköre: *Hogyan működnek a gépi tanulás megvalósításai és az ehhez kapcsolódó egyéb technológiák, van-e ezek között olyan tényező, melyet az MI-fogalomban is figyelembe kellene venni?*

K3: Az autonómia kérdésköre: *Hogyan ragadható meg az autonómia gépi és emberi megvalósulása, mik a főbb jellemzőik és a főbb különbségeik?*

K4: A kihívások és a fejlődés kérdésköre: *Milyen kihívásokat hoz a neurális MI, hogyan kezelhető a hibázása, és milyen tényezők befolyásolhatják a technológia fejlődését vagy terjedését?*

K5: A terminológia kérdésköre: *Milyen esetekben vetődik fel, hogy hiányos vagy zavaros az MI-fogalom használata, hogyan tehető javaslat az MI-terminológia javítására, és hat-e ez a klasszikustól eltérő adatkezelési paradigma „informatika” kifejezésére?*

K6: Védelmi kérdéskör: *Hogyan (milyen modellek, fogalmak, összevetések, elemzések mentén) ragadhatóak meg azok az újdonságok, melyek az MI hatására a védelmi szegmensben várhatóak?*

Választott kutatási módszerek

A fenti kérdéskörök, korábban kijelölt szempontok és lehatárolások alapján, továbbá figyelembe véve, hogy a vizsgálatok különböző fázisaiban eltérő módszerek alkalmazása indokolt, a kutatást az alábbi módszerekkel végeztem el:

Deduktív módszerekre és analitikus megközelítésre volt érdemes támaszkodnom az alapvizsgálatoknál. A dokumentum- és forráselemzés módszereit egy-egy kérdéskör eddigi kutatási eredményeinek feltárásakor vagy hivatalos dokumentumok recenziójánál használtam. A terminológiai vizsgálatoknál szerephez jutott az etimológia is, hiszen az átvett régi szavak esetén sokat elárulnak az adott szavak ősibb jelentésrétegei. Az analízisekben feltárt információk rendszerezésére összefüggés-feltáró módszereket alkalmaztam: pl. párhuzamba vagy oppozícióba állítottam a kapott anyagot, más esetekben a vizsgálat tárgyát szintekre, vagy részekre osztottam a jobb vizsgálhatóság érdekében, vagy táblázatokban tettem áttekinthetővé őket – ezek a felosztások jelen használatukon túl is alkalmazhatóak. Ám a rendszerezett, de analitikus anyag szintézisére is szükség volt, induktív módszerek bevetése nélkül nehezen jönne létre eredmény. A következtetéseket sokszor diszkurzívan, a feltárt információk és saját gondolatok logikus kapcsolataként kaptam meg, máskor az információrendezés eredményeiből vagy az összehasonlításokból is vontam le 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 is, melyekre előrejelzések vagy javaslattételek alapulnak. A könnyebb érthetőség érdekében, valamint a gondolatmenet ellenőrzésére gondolattérképeket is rajzoltam, melyből kettőt digitálisan is elkészítettem. Az így kapott gondolatsorok közlésére leíró módszert használtam, de a jelenségek és fogalmak esetében inkább visszatekintő elemzést, a folyamatok vizsgálatához pedig feltáró módszert alkalmaztam. Szükség esetén ábrákkal illusztráltam a magyarázatokat. A logikai gördülékenység érdekében alkalmaztam egy „láthatatlan módszert” is: limitált idejű vetített előadásokba konvertálva egy-egy gondolatsort, sok esetben sikerült javítani a felépítéseken, a megfogalmazásokon és a tömörségen.

Munkám során számos MI-alapú szövegfeldolgozó rendszert teszteltem, ezeket minőségi szakirodalmak felkutatására (google scholar helyett), vagy a felmerült ismeretlen fogalmak gyors leírására (wikipedia helyett) használtam. Bekezdések megírását nem tudtam az MI-re bízni, mivel az eredménnyel több utómunka volt, mintha magam készítettem volna őket (kivéve a szakirodalmi áttekintést). A gépszöveg elnyomta a saját megközelítésem sajátosságait, személytelenné téve a szöveg stílusát.

Áttekintés a szakirodalomról

Először néhány formai megjegyzés. A szakirodalmi hivatkozások mindig a főszövegben közölt információkhoz kapcsolódnak, de olyan esetekben a lábjegyzetbe kerültek, amikor magyarázó megjegyzéseket fűzök hozzájuk, vagy ott közlöm a tárgyhoz tartozó példát, idézetet vagy kommentárt – ezt nem lehet más módon áttekinthetően megoldani. A bibliográfiát egy célprogram (Zotero) segítségével hoztam létre. Sok egyetem elkészíttette a saját elvárt irodalomjegyzékét megadó formátum leíró fájljait, ám intézményünk egyelőre adós ezzel, ezért a beépített lehetőségek közül egy helytakarékos formátumot választottam, amelynek megjelenítési beállításain kézileg átprogramozva javítottam[1].

A források minőségileg megfelelnek a jelenlegi tudományos szakirodalmi elvárásoknak, de emellett a jövő kívánalmai szerint is értékeltem a szakirodalmakat. Hiszen jelen tanulmány is rámutat, hogy MI által létrehozott változások a jövőben mindenre kihatnak: azokra az elvárásrendszerekre is, melyek előírják a tudományosság kritériumait, az idézés formátumát, vagy pl. jelen alfejezet megfogalmazását. (A MI tudományossági szempontjairól ld. IV.2.5.) Az olvasó számára is evidens, hogy egy ilyen tanulmány-alfejezetnél sokkal naprakészebb, pontosabb és jobb áttekintést adnak egy témáról a nyelvi modellek. Ezért úgy döntöttem, hogy az egyik ilyen szolgáltatásra bízom az elvárt áttekintést, hiszen ez nem kreatív tudományos munka, sőt épp a gépeknek való feladat; bármely nagy modell nálam sokkal jobban végrehajtja az általam átolvasott számtalan cikk és könyvrészlet összefoglalását. Ehhez a Perplexity szolgáltatásba feltöltöttem a már elkészült irodalomjegyzék több mint háromszáz tételét, kértem tőle egy áttekintést[2], amire az alábbi szöveget kaptam. Ezt én idézetként kezelem, jelezve, hogy csak ennél a fejezetnél írattam szöveget az MI-vel.

„*A mesterséges intelligencia (MI) kutatása az elmúlt évtizedben soha nem látott dinamikával bővült, különösen a katonai és védelmi alkalmazások területén. A témakör elméleti fundamentumait olyan klasszikus művek alapozták meg, mint Turing úttörő dolgozata a gépi intelligencia lehetőségeiről, illetve a MI történeti és filozófiai vonatkozásait bemutató Buzás, Pokol és Rus-*

[1] Az IEEE formátum módosítása így pl. a teljes keresztnevet és az URL-t is kiírja, de pár formázást is módosítottam. Az program által használt leíró nyelv (CSL) bizonyos karakterek módosítását vagy törlését nem teszi lehetővé (a dokumentáció szerint sem), ezért kézileg is formáztam a kapott eredményt.

[2] Nem szószerinti az idézet, mivel a gépi választ picit javítani kellett (angol töredékszavak, fölösleges 4-5 vessző a sorvégeken, vagy jómagam a jeles kutatók között, stb.). A sablonos „összefoglalást” töröltem. A következő (igen egyszerű) promptot használtam: „A csatolt fájl egy irodalomjegyzéket tartalmaz. Készítsd el belőle egy tudományos munka "irodalmi áttekintés" című fejezetét kb. 4000 karakter terjedelemben”.

sell és társai munkái. Ezekben az alapművekben világosan kirajzolódik, hogy az MI, mint tudományterület, kezdettől fogva két fő irány között oszcillál: az emberi intelligencia utánzására (strong AI) és a szűk, specializált alkalmazások (narrow AI) megvalósítására törekvő vonulat között.

A technológia fejlődésével párhuzamosan a szakirodalom egyre hangsúlyosabban foglalkozik az MI alkalmazási potenciáljával, különösen a védelem és biztonságpolitika terén. A nemzetközi környezetben (pl. NATO, EU) több stratégiai dokumentum és jelentés készült az MI védelmi integrációjáról, illetve annak biztonságpolitikai kihívásairól. Ezek a jelentések egyrészt felmérik a mesterséges intelligenciával támogatott rendszerek előnyeit a hadviselés, a stratégiai döntéshozatal és a hírszerzés területén, másrészt rámutatnak a technológia jelentette új fenyegetésekre, az adaptív ellenfelek megjelenésére és a jogi-etikai dilemmákra. Érdemes kiemelni, hogy a NATO és az EU egyaránt nagy hangsúlyt fektetnek a szabályozási környezet folyamatos fejlesztésére, a transzparencia és bizalomépítés jegyében.

A magyar szakirodalomban is megjelentek az MI hazai stratégiáit, fejlődési irányait összefoglaló dokumentumok, amelyek illeszkednek a nemzetközi trendekhez, de régió-specifikus kihívásokat is bemutatnak. Ezek alapján hazánk jelentős erőfeszítéseket tesz az MI haditechnikai alkalmazásában és a hibrid fenyegetések elleni védekezésben.

Az MI technológiai fejlődésének egyik fő pillérét a gépi tanulás, azon belül a mélytanulás adja. A neurális hálók, a konvolúciós és rekurrens architektúrák, valamint az attention mechanizmusok forradalmasították az adatfeldolgozást és az autonóm rendszerek képességeit. Ezeknek az algoritmusoknak a jelentősége különösen a képfeldolgozásban, autonóm járműirányításban és a természetes nyelvfeldolgozásban mutatkozik meg. A katonai alkalmazásokban e módszerek lehetővé teszik a valós idejű helyzetértékelést, célazonosítást, valamint a harctéri információk gyors és önálló feldolgozását.

Az MI rendszerek autonómiája és a döntéshozatali szintek kérdése kiemelt szakmai vita tárgyát képezik. Míg a szűk MI (narrow AI) már napjainkban is kritikus szerepet tölt be, a szakirodalom komoly figyelmet fordít a mesterséges általános intelligencia (AGI) és a szuperintelligencia (ASI) potenciális eljövetelére, melyek stratégiai szempontból is paradigmaváltást hozhatnak. A szimulációk, elméleti modellek és jövőkutató munkák mellett (pl. Bostrom, Tegmark, illetve a friss iparági előrejelzések) megjelennek a vészforgatókönyvekkel, a szabályozási és etikai kihívásokkal foglalkozó elemzések is.

A MI emberközeli, "affektív" dimenziói – mint az érzelemfelismerés, empátia és az ember-gép interakció – szintén előtérbe kerültek a legfrissebb tudományos diskurzusokban. E téren

Picard, Ekman, Asada és társai munkái említhetők, melyek a gépi érzelemdetektálás, az affektív számítástechnika, illetve ezek katonai vonatkozásait vizsgálják.

A szakirodalomban külön fejezetet képeznek a természetes nyelvfeldolgozás (NLP) katonai szerepéről szóló tanulmányok, illetve az ipar 4.0, intelligens gyárak és az adattárolás-elemzés biztonsági kihívásait tárgyaló cikkek. Ezek rávilágítanak arra, hogy az MI nemcsak a klasszikus fegyverrendszereket, de az ellátási láncokat, a kritikus infrastruktúrát és a kibervédelmet is átalakítja, folyamatosan bővítve a "digitális harctér" fogalmát. E sokszínű szakirodalmi alap biztosítja a további multidiszciplináris MI-kutatások megalapozottságát a védelmi szektorban."

I. EMBERI ÉS GÉPI INTELLIGENCIÁK

Bár az MI-ről a legtöbb embernek van bizonyos fogalma, fontos, hogy a tanulmány erről szóló vizsgálatai előtt pontosítsam, mit szokás érteni alatta (I.2). Ezért szükséges a munkát a K1 kérdéskör elemzésével kezdeni, és ezzel a további kutatáshoz alapot teremteni. Azt is szem előtt kell tartani, hogy az itt leírtakhoz szervesen kapcsolódni tudjon majd K4 kérdéskör kutatása, vagyis már itt lépéseket tegyek a C1 cél felé, sőt lehetőség szerint a másik két cél irányában is. Ezeknek az elvárásoknak megfelelően mutatom be a fogalom történetét, használatát és az ágens fogalmát (I.2.), majd azt, hogy hogyan értelmezték és értelmezik a MI-t, milyen fajtái vannak az emberi intelligenciának, és hogyan szokás elképzelni az MI fejlettebb szintjeit (I.3). Ezt követi egy példa annak illusztrálására, hogy az intelligencia kifejezés nem csak az okossággal és az IQ-val jellemezhető: ehhez bemutatom, hogy az érzelmi intelligenciát hogyan képesek utánozni az affektív számítástechnika eredményei (I.4.). Először azonban rövid magyarázatot adok az első három fejezet belső összefüggéseire, és tárgyalási sorrendjükre (I.1.).

I.1. AZ INTELLIGENCIA-TANULÁS-AUTONÓMIA ÖSSZEFÜGGÉS

A tanulás, az intelligencia és az autonómia az emberek és a gépek világában egyaránt szorosan összefügg, és egymásra épül.[3]

Az élőlények fennmaradását tanulási képességeik biztosítják, a különböző szintű és típusú intelligenciáik ennek a képességnek a kibontakozásai. Az állatok és az ember intelligenciájuk alapján képesek a véletlentől eltérő viselkedés valamely szintjére, a különböző ösztöneik közötti döntésre, – az ember esetében ezeken túl akár különböző gondolatmenetek összevetésére is. Tehát az élőlényeknek 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.[4] 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.

[3] Előzetesen ezek rövid meghatározása. *Tanulás*: képes viselkedését célszerűen és megismételhető módon változtatni. *Intelligencia*: az ember következtető és kognitív képességeinek utánozása. *Autonómia*: emberi beavatkozás nélkül képes reagálni a változó környezeti behatásokra.

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

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. fejezet). Ezután pontosítani kell azt a technológiai kört, amely jelenleg szorosan összefonódik az MI-vel, el kell kicsit merülni a gépi tanulás modelljeiben, valamint vázolni a technológiákat és az elveket (II. fejezet). Ezek ismeretében térhetek rá az autonómia vizsgálataira (III. fejezet).

I.2. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK JELENE

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

I.2.1. Az MI-fogalom megjelenése és az MI hullámai

Az MI történetének egy érdekes részét, az „MI-tél ciklusokat” egy másik összefüggésben fogom majd vizsgálni és elemezni (IV.4.2.), itt a részkutatásban maradva, csupán a fogalom kialakulásának és jelentésváltozásának ismertetése a cél. Az emberi elme gépi utánzására irányuló törekvések 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.[1] 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 névadás 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. Végül a John McCarthy által javasolt *artifitial intelligence* fogalmat elfogadták el [2] – amely aztán máig „rajta ragadt” a technológián, mint valami gyerekkori becenév. Sokáig tökéletes kifejezésnek is tűnhetett, hiszen ekkor még legtöbben 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 [3, pp. 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 ebben a döntési fák, a bool-algebra és a fuzzy logika (ld. II.3.2) voltak meghatározóak. A XX. század közepén történt bíztató 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 korszellem, ami a racionalitás győzelmében hitt, az érzelmek teljes kiiktatását a gép előnyére írta.[5] 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 erőforráselosztó, 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 elkezdtek 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. Az MI végül kb. a 2010-es évektől vált közismert tömegtermékké, és kezdett el nagyon terjedni. Azóta a bio-inspirált tanulási módszerek, például a mélytanulás,a rajintelligencia és az egyéb biotikai módszerek (ld. II.3..3.) alkalmazására összpontosítanak, és jelentős sikereket értek el az érzékelés és észlelés pontosabbá válásában. Ide tartozik „2023 slágere", a humorral és irodalmi készségekkel is rendelkező ChatGPT 4.0, vagy az autonóm járművek komplex rendszerei, és az okoskibervédelmi megoldások is.

[5] 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.

Jelenleg is érezzük, hogy ez a harmadik hullám kisebb cunamiként ömlött rá a világra, gátak nélkül. A robbanásszerű, a túl hirtelen terjedés és túl gyors változás már önmagában megnehezíti a megfelelő szabályozást, nem is beszélve a problémakör újdonságáról. De nem csak jogilag, hanem a hétköznapokban is érezhetővé váltak az emberek és gépek kapcsolódásának új kihívásai, melyre a megoldást legtöbben az ún. *szimbiotikus intelligencia* megvalósulásában látják. Ezen a néven utalnak az emberi-gépi együttműködés újgenerációs lehetőségeire.[4] A hivatalos megközelítések abban látják az MI megfelelő felhasználásának zálogát, hogy sikerül az ember és az MI szimbiózisát létrehozni. Hozzáteszik, hogy ehhez az embereknek is változniuk kell majd.[3, pp. 52–53] Csakhogy 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*" szógyök miatt. A félreértést erősíti, hogy olvasni lehet ember-gép szimbiózisról is. Sokkal pontosabb lenne a szinergia (*synergia*) szó alkalmazása, 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 sikerülni fog-e a gépek és az emberek együttműködése, 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. (ld. II.3.1 – 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 (IV.) is alátámasztják az V. fejezet vezérgondolatát, miszerint nem szabad várni az MI-fogalom bővítésével és pontosításaival.

I.2.2. A mesterséges intelligencia mai hivatalos meghatározásai Magyarországon és az EU-ban

Itt a célom 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 egymás meghatározásait), ezért sem érdemes itt óriási mennyiségű MI-meghatározást összevetni csak ezekre a közös vonásokra koncentrálok. A jól használható meghatározásokat két-

oldali 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 az 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."*[6]

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".*[5, pp. 6]

Kiindulásnak, illetve alapozó oktatásnál jó definíciók ezek, de nem eléggé átfogóak. Részemről volt egy várakozás, hogy a régóta vajúdó 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 szeretnénk jellemezni, azt mondhatjuk, hogy az Európai Unióra jellemző joghangsúlyos eltolódás uralkodik a 2023 júniusában elfogadott EU-s rendelet-javaslat[7] (továbbiakban: EU AI Act [7]) MI definícióján is. A fogalmak felsorolásánál említett, agyonrövidített meghatározás sajnos inkább visszalépést jelent a korábbi EU-s meghatározásokkal szemben, vagy a világ többi részének megfogalmazásaihoz képest, hiszen nem jelennek meg sem gépi kognitív képességek, sem az autonómia problémái[8] (ezek a világ többi definíciójában általában

[6] A „mégis" kötőszó helyett egy sima és-t javasolnék (minden elismerésem és tiszteletem mellett).[5, pp. 9]

[7] A hivatalos hírek címei, 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."[6]

[8] 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, ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek.*" (3. cikk 1.)

hangsúlyt kapnak). Pedig abban a változatban, amelyet az OECD-vel is egyeztettek[9], még megjelent a virtualizáció és az autonómia is. A végül elfogadott verzióba azonban az említett vértelen meghatározás került, melyet ugyan pontosítanak a szabályzó más részében[10] felsorolt, MI-vel kapcsolatos technikák és megközelítések,[11] ez a felsorolás azonban jócskán igényelne további definíciókat. Így 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 eredményezett.[10, pp. 6][12]

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

[9] Ez a megfogalmazás még így hangzott: „*A „mesterségesintelligencia-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).[8]

[10] [9, pp. 2] I. Melléklet A mesterséges intelligenciával kapcsolatos technikák és megközelítések a 3. cikk 1. pontjának megfelelően.

[11] „*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 következtető motorokat, a(z) (szimbolikus) érvelést és a szakértői rendszereket; c) Statisztikai megközelítések, Bayes-féle becslés, keresési és optimalizálási módszerek.*”

[12] A hivatalos fordítást nyelvileg javítva idéztem.

Érthetetlen, hogy ezt miért szimplifikálták le a 4 év múlva elfogadott 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 a meghatározásokban. A felkért szakértők az *intelligencia* átfogó meghatározása helyett egyenesen az *észszerűség* fogalmát használják.[13] Ennek folyományaként logikus lenne a szakirodalomban Mesterséges Észszerűségről beszélni, bár ezt az átnevezést senki nem veti fel. És nem véletlenül nem említik, hiszen régóta zajlanak fejlesztések, melyekben az okosságon kívüli képességek gépi előállítására törekszenek. Leginkább az érzelmi (ld. I.4) vagy szociális intelligenciát [11] próbálják mesterséges eszközökkel utánozni, ezeknek a meghatározásokban meg kellene jelenniük.

I.2.3. Az USA megoldása párhuzamos definíciókkal

Az Amerikai Egyesült Államok Kongresszusának megközelítése[14] nem csupán azért érdemes a vizsgálatra, mert a világ jelenlegi vezető hatalma nyilván kihat minden egyébre. 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.

Elsősorban arra voltam kíváncsi, hogy túlmutat-e a megközelítés az imént felvetett Mesterséges Észszerűségen, megjelennek-e benne egyéb intelligenciatípusok, vagy legalább a megismerés (kogníció) fogalma, hiszen azzal is utalni lehetne a 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 egyéb intelligenciatípusok számára oszlopokat. Ahol megjelenik az adott tulajdonság, azt „X” jelöli, ahol nem, azt pedig üres mező.

[13] Az észszerűség (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*”.[10]

[14] Saját fordítás.[12, pp. 1]

[15] Itt csak röviden. A tanulás: a gép képes a viselkedését célszerűen és megismételhető módon változtatni. Az intelligencia: az ember következtető és kognitív képességeit utánozza. Az autonómia: emberi beavatkozás nélkül képes reagálni a változó környezeti behatásokra.

	A mesterséges intelligencia definíciója	autonómia	tanulás	ész (gondolk.)	megismerés
1.	Bármilyen mesterséges rendszer, amely változó és előre nem látható körülmények között hajt végre feladatokat jelentős **emberi felügyelet nélkül**, vagy amely képes **tanulni** a tapasztalatokból és javítani a teljesítményét, ha adathalmazokhoz kapcsolódik.	X	X		
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 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** feladatok megközelítésére terveztek.		X		X
5.	**Racionális** cselekvésre tervezett mesterséges rendszer, beleértve mind az **intelligens** szoftverügynököt, mind a testtel rendelkező robotot, amely észleléssel, tervezéssel, érveléssel, **tanulással**, kommunikációval, döntéshozatallal és cselekvéssel éri el a céljait.		X	X	

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

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 hiányzik ez az utolsó oszlop. Például a NATO által 2021-ben megfogalmazott MI-stratégiájának 2024-ben tervezett felülvizsgálatakor is csupán a generatív MI-t emelik ki hiányosságként.[13] De példa lehet erre a NATO

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 az iránytervben, hogy az autonómiát külön elemzi egy teljes fejezetben, ezzel azt sugallva, hogy a gépi autonómiaképesség 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. Érdemes azonban egyéb kritikai megjegyzéseket is feldolgozni, és a fenti, ötoldalú meghatározásra találtam említendő anyagot. Néhány szóban összefoglalom ennek pár fontos meglátását, melyek igazak sok hivatalos MI-definícióra is.[14] Elsősorban azon meglátása miatt említem, mely szerint a „racionális cselekvés" kifejezés indokolatlanul korlátozhatja a mesterséges intelligencia meghatározását a racionalitásra vonatkozó feltételezésekre.[17] 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. A szerzőtől nem átvettem, hiszen erre vonatkozó kutatásom kb. a cikk megjelenésével egyidős, csak később került publikálásra.[15]

I.2.4. 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 környezetünket észlelő és cselekvő ágensek tanulmányozása.*"[16, pp. xxxiv] De azért is fontos, mivel könnyű összekeverni az MI-t és az intelligens ágenst (*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*)." [3] (saját fordítás).

[17] Azzal is érvel, hogy a mesterséges intelligencia kognitív feladatokra való korlátozása korlátozhatja a mesterséges intelligencia különféle felhasználásait – ám én ebben nem látok problémát.

teszi. Tehát hasonlattal 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 igazából ez nem csupán az MI-ről mondható el, hanem még egy programozott automatára[18] is igaz. Az ágens szó első ilyen említése 1973-ban volt [17]; a szakirodalomban azóta használják, de sokáig nem az MI-vel összefüggésben, és számos típusára máig 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, folyamatszervezési 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 irat-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 ágenst valósít meg.

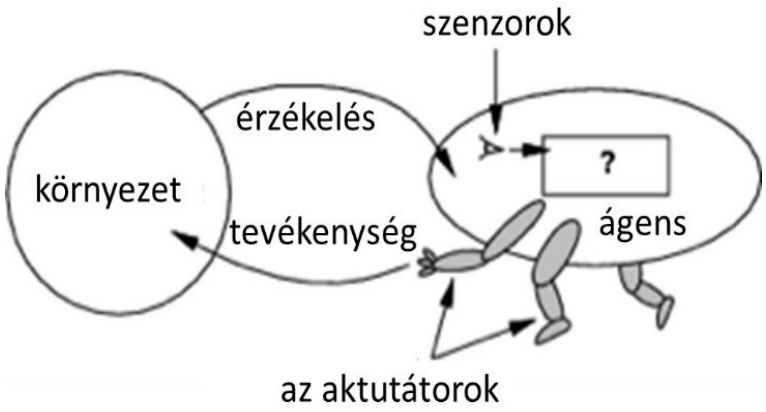
Az ágens és az MI közötti különbségnek van egy másik oldala is, melyre a fenti céges hasonlat nem világított rá, és az sem, hogy szoftverként kezeljük. Az ágens olyan, mint egy edzés, mely része valamely sportnak (ahol a sport áll analógiában az MI-rendszerrel). Az adott sport működésének része sok szabály, 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 magyarítani). 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).

[18] Itt elegendő a fogalom közismert értelmezése. Az automaták és az MI összevetéséről több vizsgálatot is folytattam, ld. IV.3, V.1.

Az 1. sz. ábra sémája alapján sokféle ágens létezik, melyeket részben a tervezésekor figyelembe vett főbb elvek, részben a felhasznált technológia (pl. tanulás), más esetekben a bonyolultság alapján különböztetünk meg. Sok szintet, illetve típust lehet elkülöníteni, [18] ám témánk szempontjából elegendőnek tartom itt csupán legfontosabbakat érinteni.

1. A legegyszerűbb szint a reflexiós ágens, mely egy környezetét érzékelő, és az érzékelt adatok alapján programozottan cselekvő automata;
2. Fejlettebb változatába beleprogramoznak egy olyan modellt is, mely segít számára a világ egy adott részének nyomonkövetésében;
3. Megkülönböztetünk cél-alapú ágenst, melyben a cél elérése kap főszerepet, ennek elérésére hoz döntéseket, és bizonyos típusai több cél között is tudnak rangsorolni;
4. Ez utóbbit fejleszti tovább a hasznosság-alapú ágens, mely egy-egy adott elérhető állapot-



1. ábra: Az ágensek működési sémája (saját átszerkesztés)

hoz mérőszámokat rendel, így nem egyetlen célt akar elérni, hanem sok tényező együttesének maximális hatékonyságára törekszik;

5. 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;
6. Egy másfajta módon fejlettebb a hasznosság-alapú ágenseknél a MAS (*Multi Agent System*), vagyis többágensű rendszer, melyben sok ágens együttműködésén van a hangsúly;
7. A többágensű rendszerek bonyolódása nyilván a részt vevő egyre több ágens tervezettebb hiearchizálását 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 visszavetítése lenne a múltba;

8. Jelenleg azok a hierarchikus ágens-rendszerek képviselik a technológia csúcsát, ahol sokféle tanuló ágens is meghatározó szerepet kap. Ezeket hívhatnánk hierarchikus tanuló / intelligens ágenseknek, ám valójában eleve csak ezt értik alattuk, így felesleges a kifejezést bonyolítani (a fogalom megértését azonban remélhetőleg segítette ez a pontosítás).
9. Végül meg kell említenünk az *embodied agent* – magyarul megtestesült ágens – kifejezést, amely alatt valamiféle külső testi jellemzőt (arcot, kezet, lábat) kapott ágenseket kell érteni. A kilencvenes évek végén[19] tűnik fel a terminus egy szoftveres, virtuális megoldásra: a társalgó robotok számára kívántak a monitoron megjelenő, mozgó arcot generálni. Később a kifejezés bővült, és egyrészt 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 fogalom mai meghatározása: *olyan intelligens ágens, mely egy fizikai vagy virtuális testen keresztül lép kölcsönhatásba a környezettel.*

I.3. 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 meglehetősen, sőt túlságosan is sokszínű képet mutatnak. Ezért itt csupán a cél szempontjából fontos aspektusokat emelem ki.

I.3.1. Nyelvi megközelítés

1. A köznyelvben általában az intelligens jelzőt használják, az intelligencia főnevet ritkán alkalmazzák; ilyenkor leginkább egy olyan sajátosságra utalnak vele, ami által valaki „*okosan és kulturáltan nyilvánul meg*”. Egyszerűsítve mondható, hogy a hétköznapokban az okos szinonimája, azonban némi többletjelentéssel bír. Egy viselkedni nem tudó zsenire inkább azt mondják, hogy „okos, de bunkó”, míg egy udvarias, figyelmes, magabiztosan 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.
2. A kifejezés latin etimológiája szerint talán az „*egybegyűjt*” szó (*inter* = közé, együvé + *legere* = szed, gyűjt)[20] adhat meg egy sokatmondó szószerinti fordítást. Ez feltárja, hogy a

[19] 2000-ben már könyv foglalkozik a témával, de sokkal korábbi publikációt nem leltem fel.[19]
[20] https://www.arcanum.com/hu/online-kiadvanyok/Lexikonok-magyar-etimologiai-szotar-F14D3/i-i-F266A/intelligens-F2767/ (Letöltve: 2024.01.15.)

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 valódi megértésük nincs.[21] A mai gépintelligencia inkább egy magolós gyermek szintjén ügyes csupán, amint erre számos kutató mutat rá.[22]

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.

1. 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 hallgat". 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.
2. Ide kapcsolódik egy másik jelentésárnyalat: a feldolgozott információ megosztása által, egyfajta kollektív intelligencia is kialakul (vö. rajintelligencia II.3.3.). Ennek a tudásnak a

[21] Az MI-nek nincs mögöttes világmodellje; megmondja, hogyan kell *hangzania* a válasznak, ám ez nem azonos a jó válasszal. Ld.[20].
[22] Az előző interjúidézet tudományos háttere: [21].

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 nyelven[23] 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 tolódik el. Pontosabban az értelmiség társadalmi szerepe és összetétele éppúgy át fog alakulni, mint számos értelmiségi dolgozó feladatköre. (Például aki csak rutinmunkákra volt alkalmas, az kieshet ebből a rétegből.)

3. Az ismeretmegosztásból adódó harmadik fontos oldal, hogy az intelligencia sok nyelvben *hír* jelentéssel, vagyis *átadott információként* is szerepel.[24] 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 régi, „értéktelen" adatot összefüggésbe hozni a kapott új információkkal, ezáltal értékessé tenni azt.
4. 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 különválik 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, tapasztalatoknak a feldolgozását, rendszerzését és okos visszaadását is jelenti. Akinek az előző mondat részeire irányuló minden képessége jó, annak az emberi közösségi dinamikák alapján

[23] Pl. a németben, angolban, oroszban az *intelligentsia* az „a szellemi foglalkozásúak közössége", azaz értelmiség jelentéssel is bír. https://uesz.nytud.hu/index.html?displaymode=web&searchmode=exact&searchstr=intelligencia&hom= (Letöltve: 2024.01.15.)
[24] Használatos a szó pl. a latinban és angolban „üzenet, értesülés, hír" értelemmel (forrás az előző jegyzetben).

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 „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.

I.3.2. Az intelligenciafajták

Az etimologizálás után ebben a kisfejezetben logikusan az intelligencia pszichológiai-filozófiai értelmezéseinek történeti alakulására kellene rátérnem. Azonban csak ismételni tudnám egy nemrégiben megjelent összefoglalás tartalmát [22, pp. 13–21], melyet átolvasva bárki arra az eredményre jut, hogy az MI fogalmához a történeti felfogás változásai általában nem adnak 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, széles körűbben 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 [23], melyet még ő maga kilencre emelt. A szemlélet alapján azonban az intelligenciatípusok száma jócskán kibővült, és a jövőben várhatóan tovább szaporodik. A mai megközelítésekben számos különféle felosztás olvasható, melyek több ponton eltérnek a korai intelligenciafajtáktól. 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.4.).

Visszatérve a sokféleségre: a különböző kutatóiskolák 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 megjelent[25] intelligenciafajtát, valamint saját meglátáson alapuló típusokat. Ez a teljesség igénye nélkül összeállított lista is több mint húsz féle intelligenciafajtát különböztet meg. A

[25] Egy kevésbé pontos és tudományos, azonban igen kreatív forrás adta az alapötleteket: [24].

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,[26] 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 típusmennyiség is óriási számú ahhoz képest, hogy a legtöbben csupán az értelmi oldalt asszociálják az MI-ben szereplő intelligencia szóhoz.

2. táblázat: 22 intelligenciatípus (több modell alapján saját összeállítás)

	Gardneri intelligencia típusok	**későbbi intelligencia fajták**	Rövidítés (ha van)
1.	logikai	intelligenciahányados	Intelligence Quotient (IQ)
2.	egzisztenciális		
3.	intraperszonális		
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 szembenézés hányadosa	Adversity Quotient (AQ)
12.		pozitívnak látási hányados	Positive Quotient (PQ)
13.		kreatív intelligencia	Creative Intelligence (CQ)
14.		gazdasági érzék hányadosa	Financial Quotient (FQ)
15.		spirituális hányados	Spiritual Quotient (SPQ)
16.		tapasztalati hányados	Experience Quotient (XQ)

[26] Pl: a *táplálkozási hányados* az étkezési (testi) tudatosságot és önfegyelmet méri.
A *viszontagsági hányados* a problémákkal való szembenézés 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 szabályalkalmazási készséget méri. Ezek összevonására nem találtam a forrásban magyarázatot, ezért említem külön sorban őket.

	Gardneri intelligencia típusok	**későbbi intelligencia fajták**	Rövidítés (ha van)
17.		digitális hányados	Digital Quotient (DQ)
18.		technikai kompetenciák	Technical and Ethical Quotient (TEQ)
19.		etikai kompetenciák hányadosa	
20.		szintézisteremtő intelligencia[27]	
21.		gyakorlatiassági hányados[28]	
22.		emberismereti hányados[29]	
23.		…	

Az intelligenciafajták 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 meglévő alapképességek a megoszlása egyénenként erősen eltérő. Ezek együtt határozzák meg az egyének mindennapjait és szakmai kibontakozását. A táblázatban is helyet kaptak (saját típusként), mivel az analizáló és szintetizáló MI-rendszerek modell szinten is el kell, hogy térjenek. A gyakorlatiasság emberi érzéke pedig erősebben szocializáció-, illetve tanulásfüggő, mint a másik két elem, – így ennek gépi megvalósítása várhatóan inkább a tanulási adathalmazok függvénye lesz, kevésbé modellfüggő.

Itt kell említenem egy másik felosztást, mely azért nem került intelligenciafajtaként a táblázatba, mivel jelentőségét nem az MI, hanem a humánvirtualizáció (saját kifejezés, ld. IV.2.2.(1))) szempontjából látom fontosnak. Ez az emberek *humán* és *reál* beállítottság szerinti felosztása. Bár felvethető ez a felosztás az MI-rendszerekkel összefüggésben is, de célrendszerekkel kapcsolatban nem találtam jelentős szempontnak (az ún. általános MI-nél jöhet majd szóba, ld. alább).

[27] Saját felvetés, a tudományos mérhetőség metódusai nélkül. Egy olyan elvonatkoztatási képességet értek alatta, amikor valakiben felébreszthető a motiváció az általa ismert dolgok szélesebb összefüggésbe helyezésére, és ezzel képes is megvalósítani információk újszerű szintézisét.

[28] 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.

[29] 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. A hétköznapokban pedig nem kell csalódnia folyton a másik emberben.

I.3.3. A gépi intelligencia tudásszintjei

Az előző alfejezethez való szoros kapcsolata miatt logikailag itt érdemes bemutatom az MI tudásszintek szerinti felosztását (számos egyéb felosztás: V.4.1.-ben). 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, enélkül 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.[16, pp. 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 háromszintű felosztás elterjedt, [30] amihez hozzá kell tenni egy félig technikai kifejezést is

1) ANI[31] – Mesterséges Vékony Intelligencia (MVI): az embernél gyengébb képességű, feladatorientált gép;
2) AGI[32] – Mesterséges Általános Intelligencia (MÁI): az emberhez hasonló képességű, általános célú gondolkodógép;
3) ASI[33] – Mesterséges Szuper Intelligencia (MSzI[34]): az embert meghaladó képességekkel rendelkező gép, mely talán az evolúció következő szintje;

\+ *technológiai szingularitás* – az ember és gép közvetlen kapcsolódása által a gondolkodás és egyéb képességeink beláthatatlan kiterjesztése.

Az ANI tehát a korábbi felosztás *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 célspecifikus rendszerektől a sokmodulos MI-architektúrákig. Tehát egy szokásokat megtanuló

[30] A három szint közérthető összevetése: [25].
[31] ANI – Artifitial Narrow Intelligence, Vékony MI, (a mai technológiák).
[32] AGI – Artifitial General Intelligence, Általános MI.
[33] ASI – Artifitial Super Intelligence, Szuper MI.
[34] A magyar rövidítések sajátok, ismeretlenségük miatt jobbnak ítéltem helyettük az angol verziókat használni.

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 futurológiai oldalról értelmezhető.

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

Az ASI lehetne simán Szuperintelligencia[35], 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 ember számára elképzelhetetlen, sőt valószínűleg felfoghatatlan képességeket, irányokat valósítana meg.[36] 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.[37]

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.[29] A nevét onnan kapta, hogy a fekete lyukak közelében tapasztalható gravitációs szingularitás értelmezhetetlenné teszi a fizika megismert szabályait. Hasonlóképen teszi értelmezhetetlenné az eddigi emberi kultúrát az emberi képességek vizionált kiterjesztése, mely által az elme közvetlenül, a testi kommunikációs mechanizmusok (kéz, hang, érzékszervek) megkerülésével vehetne igénybe számítógépes szolgáltatásokat.[38] Ez tehát egy szélsőséges diszrupció, ami nem csupán az ipart forgatja fel, hanem az ember önértelmezését is. Az elképzelés szerint az agykérget összekapcsolva a felhővel a gondolkodásunkat emelhetjük új dimenzióba, ezáltal

[35] Egyébként a névadó is csak így hívja a teóriát taglaló könyvében: [26].
[36] Más megközelítésben új létréteg jönne létre ld. [27].
[37] Közérthető áttekintés az ASI-ról: [28].
[38] Egy egész fejezetet szentel ennek pl. [30].

az MI képességeit saját képességünkként használhatnánk.[39] E szerint a testi képességek kitágításával, a szaporodás és a halál eltörlésével (mesterséges test által) az élet fogalma[40] is megváltozhatna. Az ember kiterjesztésére sok irányban zajló törekvésekről említés szintjén túl nem térek ki.[41]

I.4. ÉRZELMI INTELLIGENCIA UTÁNZÁSA A GÉPEKBEN

A fentebb (I.2.2.) felsorolt számos intelligenciatípus 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 „ezüstérmes" 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 érzelmek gépi érzékelése és utánzása; vagyis növekszik a kereslet, tehát erős a fejlesztői motiváció is, így ez a terület is igen dinamikusan fejlődik. Azonban, – bár eredményei sokszor látványosabbak, mint a gépek „okossága", – mégis kevéssé közismert, hogy ez a terület külön fejlődik, az érzelmek 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 érzelmeket 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 érzelmeket. Érdemes azzal kezdeni, hogy bemutatom az érzelmek sokrétű felértékelődését a társadalomban, így a technológia mögötti motivációk is megragadhatóbbá válnak.

I.4.1. Háttér a technológia mögött: az érzelmek fontosságának története dióhéjban

Európa történelmén jól megfigyelhető egy olyan tendencia, hogy az intellektuálitást előtérbe helyező korszakok után az érzelmeket is felfedezik újra és újra. Ahogyan a felvilágosodás racionalizmusát a romantika érzelemközpontúsága követte, úgy a II. világháború utáni modern ra-

[39] Kurzweil nem csupán kitart korábbi jóslata mellett, hanem legújabb könyve szerint ez már a küszöbön van.[31]
[40] Az élet fogalmának újragondolása a technológia tárgyalt szintjei mentén már jó néhány éve [32] elkezdődött.
[41] Egy szakdolgozatban kaptam erről átfogó képet, mely alapján világos, hogy a téma összetettsége miatt említésen túl nem foglalkozhatok vele.[33]

cionalista korszak értelemközpontú világa is lassan szertefoszlott, fokozatosan egy posztmodern korszaknak adva át helyét, ahol az érzelmet mind tudományosan, mind pedig 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 területet, bár sokszor tűnt úgy, hogy sikerül az értelmet és az 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, és azzal a céllal, hogy rámutassak, hogy már a 19. század végén megérett az idő az újszerű érzés-elmélet teóriák megfogalmazódására. Akkor egymástól függetlenül jelentetett 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.[34, pp. 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 kiválthatóak. 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 érzelemközpontúságának hajnalán jelent meg. Célja nem érzelemkezelő gépek megalkothatósága, hanem az ember érzelmi oldalának megragadhatósága volt. Azt a paradigmaváltást az érzelmek kutatásában, melyet később az elektronika felhasználhatott, Silvan S. Tomkins amerikai pszichológus és teoretikus affektuselmélete[42] 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 em-

[42] Magyarul a „hatáselmélet" kifejezést is használják. Itt az idegen szó megtartása a később bemutatandó affektív számítástechnikával való kapcsolat miatt indokolt. Vö. [35].

beri 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 alapaffektusba rendszerezte és háromféle előjelű osztályokba sorolta azokat a sémákat, melyeket a másik emberen felismerünk:[43]

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

Később különböző számú affektussal dolgozó sémák születnek, és több tudományos elmélet is létezik az alapérzelmek csoportosítására.[44]

A pszichológiában az affektív jelenségek átfedéseiként választják szét az érzelmeket (*emotion*), és ennek konkrét, szubjektív eseteit, ahol megkülönböztetik az érzést (*feeling*), a hangulatot (*mood*) és az attitűdöt (*attitude*).[45] Jelen megközelítést feleslegesen megbonyolítaná, 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 affektuselmélete 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.[46]

A script-elmélet szerint az emberi viselkedés leírható minták és érzelemmintázatok segítségével. A gondolkodásminták 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 érzelmeknél bemutattam, csak kisebb léptékű. A számítástechnika hőskorszakában 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 gondolkodásmintáinak segítségével sokkal hatékonyabb MI-modell tervezhető. Ez a mintaalapú matematikai modell mára jól bevált technológiává fejlődött, sőt a jövőre nézve is joggal gondolhatjuk,

[43] Ehhez jó ábra is készült: [36, pp. 68].
[44] További csoportosítások: [37, pp. 293].
[45] Elsősorban ezen kötet *Nyelv, tudat, gondolkodás* c. részfejezete: [37, pp. 290].
[46] Bár ezen terület alapműve Tomkins tanítványa nevéhez fűződik: [38].

hogy az MI fejlődése erre alapulva halad tovább. Itt sajnos csupán említeni tudtam, mivel az affektuselméletnek 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 hippi mozgalomban és a szexuális forradalom légkörében jogi szinten és hétköznapokban 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 örömöt az álmosolytól megkülönböztetni.[47]A nyolcvanas évekre így a tudomány szintjén is népszerűvé válik az érzelmek felértékelése, a mérhetőségének és működési modelljeinek csiszolása. Ekkortájt válik le az érzelmek mérése az intelligenciával kapcsolatos vonalról, elsősorban a fentebb említett (I.2.2.) többszörös intelligenciaelméletre alapozva.[23] Ekkor kezd jobban kibontakozni az affektuselmélet is, és vele az affektív tudomány.[40] 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 [41] 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 médiamanipuláció 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.

I.4.2. 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.[42] Számos, később megvalósuló technoló-

[47] Újabb kiadásban ld. [39].

giát megjósol ebben az írásában – a hordozható számítógépektől (nála *affective wearable computers*) kezdve egészen a médiafelhasználásokig. Rámutat arra is, hogy a Turing-teszt[48] 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, mely a kutatónő 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 intonációfelismerő [44, pp. 1097–1108] vagy kézmozdulat-felismerő [45] 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 magyarítható úgy, hogy „érzelmekkel operáló számítástechnika”[49]. 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.*[46]

Fontos hangsúlyozni, hogy a „gépi érzelem” éppen annyira nem érzelem, 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 iránynak, 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 komputer

[48] A teszt azzal méri a gépek képességeit, hogy a tesztelő meg tudja-e különböztetni, hogy kérdéseire gép vagy ember válszolt. Az eredeti teszt: [43].
[49] Online kézikönyvekben felbukkan az „érzelmi számítások” és az „affektív számítások” kifejezés is.

(output affektusok). Így négyféle géposztály létezhet – ezt példázzák napjaink jelentősebb eredményei 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 érzelemkifejezést, 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 érzelem" (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 Érzelem) vagy az angol *AIE system* (*Artifitial 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.[47] Nem pontos a szóhasználat, hiszen a technológia lényege nem kifejezetten az érzelmek kifejezése, hanem a gyorsabb és jobb grafikai ábrázolás volt. De előremutató 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, és egészen 2012-ig gyártották a különféle EE-chipeket.

Fogalmilag egyébként felvethető, hogy ennek mintájára a mai, még finomabb érzelemkifejezésre alkalmas szoftvereket, sokmagos videokártyákat és egyéb eszközöket is az affektív számítástechnikához kellene 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.[48] 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 mai eredményeinek minden kezdeménye. Megemlítendő itt, hogy Marvin Minsky figyelme is az mesterséges érzelemkutatás felé terelődött [49] (az ő neve fontos az MI történetében, hiszen számos fejlesztés és találmány kapcsolódik hozzá). 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 felhőtechnológia 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 beszédértés, és magyarul is egyre kevésbé gépies a beszédszintetizálás.

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

A 2005 utáni korszakra jellemző még, hogy az audio- vagy videoszenzorok megújulásán túl teljesen újszerű érzékelők sora jelenik meg: kezdve a Nintendo cég Wii mozgásérzékelő játékkonzoljától (2006)[50], a fizikálisan kapcsolódó szenzorokkal végrehajtott érzelem-felismerésen át (szemmozgáskövető szemüvegek, EEG-távirányítók[51]) egészen a mikrohullámú testradarokig (ld. következő szakasz). A játékok mellett az érzelemfelismerők húzóágazatává kezd válni például az autóipar is a fáradtság-ellenőrzőkkel, a vállalati munkatársjólléti alkalmazások[52], de a katonai kutatások szintén segítik a terület fejlődését.

[50] Ez egy kézben tartható, de vezeték nélküli alkatrész *(remote)* mozgását detektálta infravörös érzékelővel.
[51] Elektroenkefalográfiás fejérzékelő: nem hagyományos effektust érzékel, hanem elektrofiziológiai változásokat, az idegsejtek elektromos aktivitását méri valós időben.
[52] Ez már 2005-ben prognosztizálható volt olyan ötletekkel együtt, mint a csoportdöntési segédeszközként való használat vezetők számára, hogy feltérképezhessék a beosztott állományuk hangulatát. Ld. [50, pp. 97].

Ezek közül azt a török katonai kutatást szeretném röviden bemutatni [51], amely a hagyományos neuronháló érzelmi bővítését vetette fel a célpontfelismerési képességek javítására. Dr. Khashman kutatócsoportja 2009-ben egy olyan megerősítéses tanulási modellt alkotott, melybe 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 a 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 érzelemmodellek, 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.[52] 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 érzelemre 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 koncentrálva sikerül megoldani számos kihívást. Ezek az eredmények később gyakran beleolvadnak nagyobb rendszerekbe, ez esetben például ma már az MI-rendszerek empátiára tanításáról beszélnek.[53]

(3.) Érzelem-chip, affektus-radar, mesterséges empátia

Az utóbbi évtizedben ugyan az MÉ2 ugrásszerűen 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 nagyvállalat[53] üzemeltet.

Az alábbi eredmények előzményei a XXI. század elejére nyúlnak vissza. A cég vezetője 2005-ben egy EU-s díjat[54] nyert el a 3D Solar technológiáért. Ez nem más, mint a 3D megjelenítés tökéletesítése, mely az Emotion Engine 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. Háromféle kiadása létezett.[54]

1. egy felhőalapú szolgáltatás (mely a megvásárlás után a mai napig működik);
2. egy beszerelhető mikrochip, amit eMCU-nak[55] hívott a cég (alább bemutatom);
3. valamint egy USB-csatlakozású perifériaként forgalmazott termék, mely kész okoseszközöket volt hivatva érzelmi komponenssel kiegészíteni.

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

Rugalmassága mellett az érzelemkezelés eddigieknél sokkal nagyobb pontossága, valamint az input- és outputképességek együttes jelenléte tűnt bíztatónak. 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 project a fentebb bemutatott Ekman-féle modell módosításán[56] alapul. Működése hasonlít egy színes szkenner működéséhez, ahol egy kép adott pontjának színárnyalata a három alapszín különböző intenzitású keveréséből áll elő. Csak itt a három alapszín erősségének értékei helyett a modell 12 alapérzelmének[57] különböző intenzitásából „keveri ki" (hozza létre) a rendszer az affektus-

[53] 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 az EPU felhőalapú része elérhető.

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

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

[56] A szakirodalom szerint a félelmet, a dühöt, az undort, a szomorúságot, az örömöt, a meglepődést és az érdeklődést a legtöbben alapérzelemnek tekintik, ám ezek közül csupán hat szerepel az EPU-ben. vö. [37, pp. 293]

[57] A 12 itteni érzelem 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.

pixel[58] konkrét árnyalatát, mely az adott pillanat érzelmét pontosan definiálja. Több időintervallumban felvett érzelem-pixelek sorát feldolgozva alakítja ki rendszer azt az érzelemmátrixot, mely számunkra információval bírhat. Ugyanis az érzelmek inkább egy időintervallumban 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 érzelemárnyalatot kezelhet [54], 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. arcgenerálással, nyelvi modellel) is képes kommunikálni. Így valósulhat meg az MI és a robotok spontán (nem kódvezérelt) érzelme, olyan nevetése vagy sírása, mely hiába csak szimuláció és csak virtuális, mégis egy eddig nagyon hiányzó problémát oldana meg.

Az audio és 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 érzelemfelismerő 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)[59] névre hallgató szenzort fejlesztettek, amely vezetéknélküli (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 NLP[60]-rendszerekhez történő integrálás is a célok között volt, azaz élethűen beszélgető gépek, illetve NLG-algoritmusok[61] 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

[58] Saját kifejezés (!), ez a leírásokban nem szerepel.
[59] Amint az imént említettem, az aRadar nevű kutatás a cég megvásárlása miatt jelenleg nem elérhető.
[60] Natural Language Procession, természetes nyelvfeldolgozás.
[61] Natural Language Generation, természetes nyelv generálása. Ez némileg eltér az NLP-től, mivel inkább adatok értelmes szöveggé alakításáról szól, ld. [55].

reagálni a felmerülő szituációkra, így téve élethűbbé a játékélményt.[62] Ezen technológiák utódairól ugyan nem leltem fel adatokat, de az említett MetaSoul csoport által feltehetően nagy szerepet játszanak a jelenlegi és várható MI-szolgáltatásokban.

Az aRadar és az általam felvetett szenzorok tehát kiegészíthetik és jócskán pontosíthatják az MÉ2 gépek által felismert érzelmek meghatározását. Az így kiegészített rendszer becsapásához nem csupán egy álmimikát 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.2.3.).

I.5. RÉSZÖSSZEFOGLALÁS: HOGYAN LEHET AZ INTELLIGENCIA MESTERSÉGES

A részkövetkeztetésekben az R utáni szám a fejezet száma, ezt követi a részkövetkeztetés száma, esetenként egy betűjel is (amikor érdemes volt egy konklúziót alpontokra osztani).

Összegzés. Az első részkutatás a tervezett egyéb vizsgálatokkal összefüggésben adott választ a K1 kérdésre, mely azt firtatta, hogy „*Hogyan értelmezik a világban a mesterséges és az emberi intelligenciát és miként fogalmazzák meg ezeket?”* A szűkebb értelemben vett választ K1-re a fogalom rövid története, a hivatalos definíciók adták. A tágabb értelemben vett válaszra csak a további kutatásokhoz kapcsolódóan volt lehetőség kitérni, ennek fényében tisztázni a mesterséges és az intelligens ágens fogalmait, etimologizálva értelmezni az intelligencia fogalmát, ismertetni az MI „szintjeit”, valamint bemutatni az emberi intelligencia fajtáit. Végül az érzelmi intelligencia gépi utánzásának bemutatásával jobban kidomborodott, hogy a gép csupán a jelenségek szintjén utánozza az embert, pl. nincsenek érzelmei csupán az azok kifejezését képes lemásolni. A leírtak egyrészt megfelelő értelmezési alapot biztosítanak a többi fejezethez, másrészt sikerült lépéseket tenni nem csupán a C1 cél felé, de C2 irányába is.

[62] https://metasoul.one/ Megtekintve: 2024.01.28.

P1-gyel kapcsolatos következtetések

R1.1. Az intelligenciafajták sokféleségére érdemes tudatosan utalni az MI minden definíciójában (I.3.2.), de jelentősége miatt az érzelmi intelligenciát külön is ajánlatos megjeleníteni benne (I.4.).

P2-vel kapcsolatos következtetések

R1.2. Nem csupán a gépi „okosság” alkalmazható támadó, megfigyelő, védekező vagy egyéb módon, hanem az gépi „érzelmek” utánzása is lehetőséget ad automatizált információs (érzelembefolyásoló) műveletekre vagy visszaélésekre (I.4.).

R1.3. A polgárok titkolt érzelmei egyelőre jogilag nem képezik személyes és intim szférájuk részét, mint pl. titkolt betegségeik, tehát ezekről a jog megsértése nélkül megszerezhetőek adatok az MI-t használva, akár visszaélési céllal is. (I.4.)

II. A TANULÁS GÉPIESÍTÉSE ÉS AZ MI-HEZ KAPCSOLÓDÓ TECHNOLÓGIÁK

Az MI-vel kapcsolatos tévhitek mögött legtöbbször a technológia alapvető nem értése vagy félreértése fedezhető fel. Ennek elkerülése érdekében C1 célhoz, egy jobb MI-fogalomhoz, elengedhetetlen legalább egy tömör válasz a K2 kérdésre[63]. Ezen belül szükséges megértetni a mai MI-rendszerek kulcsát, az „MI belsejét" azaz a mélytanulást (II.2.). Emellett szükséges az MI-hez kapcsolódó technológiák és elvek, vagyis az „MI külsejének" bemutatása is (II.1.). Továbbá egy vegyes témájú alfejezetbe szedtem össze az MI-hez kapcsolódó, és a P2 témakör vizsgálatai szempontjából fontos területeket (II.3.). (Az előző és a következő fejezet logikájához hasonlóan az emberi tanulást is elemeztem, ám ez a kutatás itt nem kaphatott helyet [56].)

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

Lassan szinte minden technológia kapcsolatba hozható az MI-vel. Itt muszáj azokat bemutatom, 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 vázolnom. 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. Adattó avagy Big Data: strukturálatlan adatok feldolgozása

A Big Data-t az 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".[57] 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. Ezért használom én is szívesebben a terjedőben lévő „adattó" (*DataLake*) kifejezést, bár elterjedtsége miatt olykor a Big Data-t is. Mivel nemzetközi vagy nemzeti szervezetektől nem leltem

[63] K2: A technológia és tanulás kérdésköre: *Hogyan működnek a gépi tanulás megvalósításai és az ehhez kapcsolódó egyéb technológiák, van-e ezek között olyan tényező, melyet az MI-fogalomban is figyelembe kellene venni?*

fel hivatalos definíciót, a félhivatalos meghatározások közül egy orvosi szervezet megfogalmazását választottam ki ismertetésül, mivel igen találóan fogalmaz [58]: „*Az adattó egy adattároló hely, amely hatalmas mennyiségű nyers adatot tárol eredeti (natív) formátumában, így ezek az adatok lehetnek strukturáltak, strukturálatlanok vagy félig strukturáltak. Ez ellentétben áll az adattárházzal, amelyben az adatok egy közös adatmodell szerint vannak strukturálva*." A teljes leírást nem ismertetem, de fontos, hogy kiemeli még a megoldás rugalmasságát is, hangsúlyozva, hogy „*a struktúra hiánya „adatmocsárhoz" vezethet, ha nem gondosan kezelik*". hasonlóan, az idézett Big Data definícióban is 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 karaktersor-keresést fájlok tartalmában (nem a fájlnévben), és pár ezer fájl átnézésének hosszú perceit vessük össze azzal, amikor egy több tízezer cellás excelfájlban 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 humánerős 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 halmazá*-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 már a kapcsolatokat is ábrázolták azzal, 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 redundáns[64] módszerré válnak. A redundancia elkerülésére sajátos matematikát és számos algoritmust dolgoztak ki, ezek által már az 1970-es évektől megvalósultak a relációs[65] elektronikus adatbázisok, melyek máig egyfajta megoldást jelentenek. Ezek számos kis adattábla irányított összekapcsolásával nagyságrendekkel hatékonyabbak a digitalizált táblázatoknál, és hatékonyan lekérdezhetőek (például SQL[66] lekérdező nyelven).

[64] Az ismétlődő adatokat nevezzük redundásnak, 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 adatábislekérdezések írására kifejlesztett kódcsalád.

A gondot itt az igen jelentős humánerőforrás jelentette. Már egy digitalizált táblázatalapú (pl. excel) feldolgozáshoz is szakértelem kell. De sokkal jelentősebb szakértői erőforrások kellenek 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 címkékkel lássa el, a könnyebb megtalálás érdekében.

A manapság termelődő adatmennyiség ilyen strukturált rögzítéséhez és feldolgozásához a humánerő 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 fájltípusaiig. Az IoT eszközök megsokszorozzák ezt az adatmennyiséget, és csak bizonyos címkéket lehet automatikusan létrehozni. Pl. egy térfigyelő kamera esetén adott a szenzor helye vagy a felvétel időpontja – ám ez ember vagy jármű keresése esetén kevés. Sőt, 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 felgyülemlő adatmennyiség egyre változatosabb. Ráadásul elég régóta keletkezik egyre több adat – így hozzá kell venni a problémákhoz a sebességtényezőt is: egyre nagyobb sebességgel 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 az adattó probléma lényegére, és a három V-hez egyre több „V" kapcsolódik.[67]

Az „adatbányászat" kifejezés szorosan kapcsolódik az adattóhoz: ez a hagyományos adatlekérdezés helyett nyit új dimenziót. Az adatbányászat nagy mennyiségű adathalmaz olyan irányított elemzési folyamata, mellyel új *információk* nyerhetőek 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 adattalálatokból kapott információkat egészíthetik ki a kapcsolattalálatok új dimenziós információi. Az ilyen mintázatfelismerést támogató a MI-rendszerek által válnak a számítógépek adatfeldolgozó rendszerekből valódi információfeldolgozó rendszerekké, hiszen az adatokból

[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]

már nem csupán az emberi értelmezés által válik információ, hanem a gép képes rámutatni az új információkra.

Ez a technológia tehát számos ponton összeszövődik az MI-vel. Már a tárolás technológiája az úgynevezett neurális adatbázis irányába halad (II.3.1.), és a lekérdezésben is alapvető szerepe van az MI-nek. A 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 adattó 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 ne használják fel, vagy bizonyítani ennek megtörténését.

II.1.2. Felhőtechnológia és technológiai virtualizáció

A meghatározás az amerikai Technológiai Szabványok Nemzeti Intézete (NIST[69]) szövege [61] jó kiindulás, mert tömör, ám meglehetősen nehézkes. A dokumentum még 5 jellemzőt is meghatároz hozzá, továbbá 3 szolgáltatási és 4 telepítési modell segítségével pontosítja a definíciót.[61] Itt ezek nem alapvetőek, ezért eltekintettem részletes ismertetésüktől (a szakma úgyis ismeri). Viszont kiemelendő az a látványos a fejlődés, amely miatt máris eggyel több szolgáltatási modellt (tehát már 4-et) tartanak nyilván.[70] A lényeg megértetése kedvéért sok forrás alapján dolgoztam át ezt a definíciót az alábbi megközelítő meghatározásokká. Először nagyon egyszerűsítve közelítem a témát, aztán picit pontosabban (és a felvetett kifejezéseket magyarázva). A digitális felhő lényege tehát 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 szállítmányozóhoz vagyunk hasonlóak. Csak mi nem vagonokat kapcsolunk össze, hanem hardveres erőforrásokat, és nem a szerelvényt osztjuk szét a

[68] Két ilyen visszaélést is bemutat: [60].
[69] National Institute of Standards and Technology.
[70] Ugyanis már 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). A teljesség kedvéért a három régebbi szolgáltatásmodell: az infrastruktúra-mint-szolgáltatás (Infrastructure as a Servie, IaaS, amikor egy virtuális gépet kap a bérlő) a szoftver-mint-szolgáltatás (Softver 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).

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.

- 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, és már a harmadik generációs gépekben (az 1970-es években) megjelent[71] 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 modernebb gépek-

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

ben 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.[72] 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 vonatos 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 erőforrás-gazdálkodáshoz 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.
- 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.[73] 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 a lényegesek. A technikai oldal pontosításához sok mindenre ki lehetne térni, itt azonban csak azt emelem ki, hogy a meghatározások általában szolgáltatásként tekintenek a felhőre, ami a szoftveres szint, pedig a fogalomba hasznos lenne beleérteni az ezt megvalósítani képes hardvertechnológiákat is. A hatékony működéshez szükséges a gépek óriási számítási teljesítménye, a robusztus tárolási kapacitása, és az óriási sebességű adatátvitel

[72] Ha akarunk különbséget, akkor a virtuális szót inkább komplex, bonyolultabb, jól paraméterezhető emulációhalmazra értik.
[73] Ez a tulajdonság függőséget és a valós világtól való menekülést is okozhat, de gyógyterápiák során is sikeresen alkalmazható (ld. affektív számítástechnika I.4.).

– mind a belső alkatrészek között, mind pedig a hálózaton. Ezek bizonyos szintje fölött kezdett csak elterjedni a technológia, ma pedig a sokmagos erős processzorok, nagy és gyors memóriamodulok, az SSD[74] tárak újabb generációi, az 5G vezetéknélküli átvitel, és minden egyéb korszerű hardver alapvető szerepet játszik népszerűségében.

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 hálózati kapcsolat nélkül, offline módban ne lehessen használni egy robotot, drónt, vagy fontos számítógépes funkciót. Ennek a 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. a mobiltelefonok 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 „közösen tanulnak"; amit az egyik gép megtanul, azt feltölti a felhőbe, ahonnan minden letöltődik a résztvevő gépekbe, és így a többieknek nem kell mindenre maguknak „rájönniük" a jobb működéshez.[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 erdejében kiigazodni, és elég keveset hibázni. A kereszteződések, táblák, lámpák felismerési nehézségeit a többiek által megtanultak alapján tudja minden autó leküzdeni.[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" (eszközök, 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, hiszen biztonságosabb egy intranet. Az IoT általam is használt hivatalos meghatározása[76] egyébként az

[74] Solid State Drive – a pendrive-oknál is használt tartós memóriatechnológia
[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]

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 időjárásszenzor amellett, hogy adatokat küld egy előrejelző rendszer számára, közvetlenül is aktiválhatja a közelében lévő optikai eszköz párátlanító 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 közeledő 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 az adattó (Big Data) 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 az 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 arról az új gazdasági ökoszisztémáról, melyhez az imént bemutatott három technológia és az MI használata együtt szolgáltattak 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

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

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 vevőkig tartó hosszú ellátási lánc egészét egyben áttekintő optimalizálásra. Ez az igény motiválta a technológiák fejlesztéseit is, hiszen multinacionális gazdasági komplexumok tudják az MI és az adattó által nyújtott előnyöket igazán meglovagolni.

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 a szűkebb megközelítésnek az előnye, hogy jobban vizsgálhatóbbá tesz számos kihívást és biztonsági problémát, amelyek az Ipar 4.0 terjedéséhez kapcsolódnak.[68] Most egy példa is elegendő szemléltetésére: az 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 említésszintű pár sor.

II.1.5. Az MIKT fogalom bevezetése

Az MI szoros kapcsolata az imént ismertetett rendszerekkel, véleményem szerint indokolttá 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 Artifitial Intelligence and Related Technologies kifejezést takarja. Tömör meghatározása 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 technológiahalmaznak a legfontosabb elemei az Adattó (más néven Big Data), az IoT, valamint a felhőtechnológia tágabb (a hardverfejlettséget is a felhőhöz soroló[79]) értelmezése.*

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

[79] A tágabb értelmezés nem szolgáltatásként tekint a felhőre, hanem abba 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.

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 technológiacsaládként 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 megközelítés, melynek lényege, hogy profitorientáltan koncentráljon a hatékonyságra. Ez azonban nem érvényes az MIKT állami vagy védelmi alkalmazásaira, ahol sokszor nem elsődleges szempont a hatékonyság: ezeknél indokolhatják a „ráfizetéses” használatot olyan tényezők, mint az emberközpontúság vagy a nemzeti érdekvédelem és hasonlók. Továbbá az Ipar 4.0 a hatékonyság érdekében az MIKT mellett számos egyéb technológiát is felhasznál, a blokklánc (blockchain) módszertől a 3D nyomtatásmódokig, melyek az MI-vel nem állnak olyan szoros kapcsolatban, mint a fentebb említett technológiák.
- 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 kézivezérlésű 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 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. energiatermelés 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

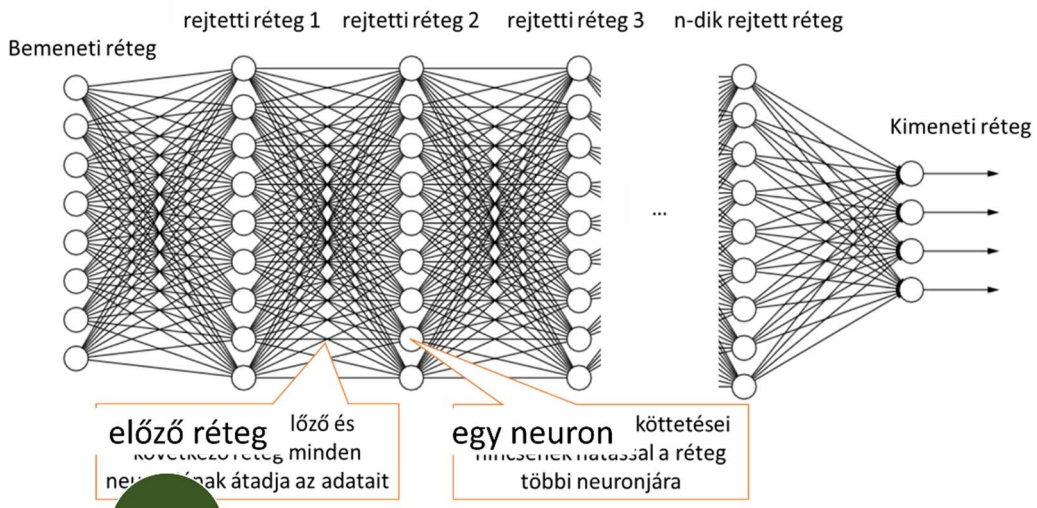
Az alfejezet célja, hogy a neuronhálós gépi tanulás fontosabb megvalósítási módjairól képet adjon, hiszen ez a technológia áll a mai értelemben vett MI-rendszerek mögött. Ezt megkísérlem röviden, de a matematika iránt kevésbé érdeklődők számára is érthető formában megtenni. Elöljáróban meg kell jegyezni, hogy tágabb értelemben a hagyományos számítástechnika is tanul, hiszen képes eltárolni, majd visszaadni adatainkat, sőt akár szokásainkat is. Az ilyen gépi „tanulás" és a mélytanulás közötti különbségre ezen alfejezet végén térek ki (és V.1.4.-ben is).

II.2.1. Az ML-modellek közös jellemzői

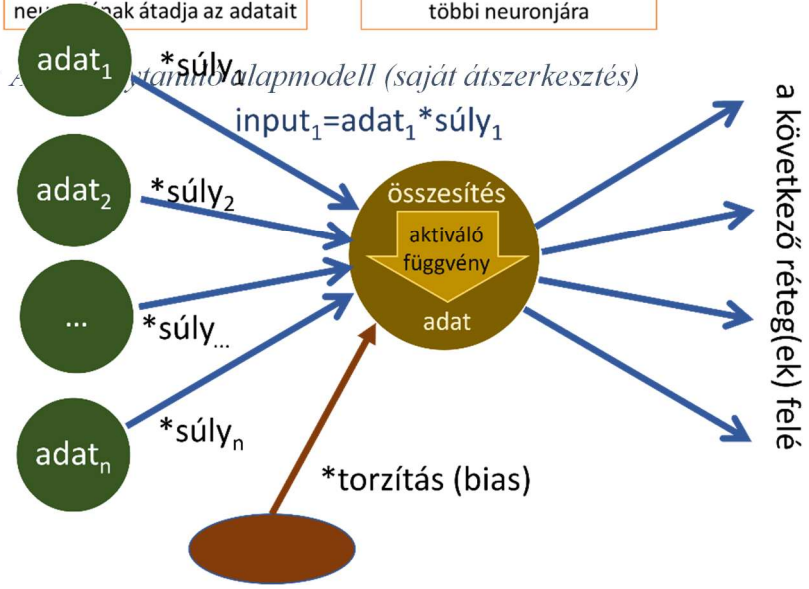
A modell alábbi bemutatása nem a műszaki ismeretekben jártas olvasók számára készült, csupán az MI működését jellemző tényezők okának ismertetése a célja. A gépi tanulás alapját az ún. „mesterséges neurális hálózat" (artificial neural network, ANN) jelenti, ami nevét az emberi agyban működő idegsejtek hálózata alapján kapta, hiszen ezt próbálták mesterségesen utánozni. Az MI egyik szimbóluma a mély neurális hálózat (Deap Neural Network, DNN), sémáját a 2. sz. ábrán[80] mutatja, ami kinézetre akár egy op-art művészi alkotás is lehetne.

[80] A szerkesztett ábrához a kiindulás: [69].

A 2. ábrán is látható, hogy a neuronok több rétegbe vannak szervezve, ezek száma adja a rendszer mélységét, ebből származik a mélytanulás (Deap Learning, DL) kifejezés. A bemenetek csak az első rétegnek adnak át információt, majd ez az első réteg 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. Valójában nem kell több belső réteg, sőt, időrendben először az egyetlen belső réteggel rendelkező változat volt működőképes Perceptron[81] néven, az 1950-es években.



2. ábra: A [illegible] alapmodell (saját átszerkesztés)



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

Nemcsak a rétegek és neuronok száma, de neuronok kapcsolata is eltérő lehet; a következő részben a fenti modellnek több ilyen variánsát bemutatom. De közös minden DL-modellben,

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

hogy az azt alkotó neuronok kimenete időben változik a tanulás során. Ezt a változást a neuronokat ért bemeneti adatok változásai idézik elő, ezért is hívják adatvezéreltnek ezt a tanulást. Ahhoz, hogy tanulás végére hasonló adathoz ugyanaz a kimenet tartozzon (pl. felismeri, hogy a képen alma van), a modellen belüli adatátvitelekhez különböző súlyértékeket rendelnek, és ezeket változtatják a tanulás során.

A rendszer alapegységeit képező neuronok valójában kicsi számítási egységek, melyek két műveletet végeznek el. Minden bemenethez külön súlyérték tartozik (ld. 3. ábra – ezeket majd tanítási lépések során számolja ki a gép). Első lépésben a neuron megszorozza ezekkel a súlyokkal a bemeneti adatokat, majd ezeket a súlyozott bemeneti adatokat összeadja. Ehhez a súlyozott összeghez még hozzáad egy saját ún. *bias* (eltolási)[82] értéket, mely magát a neuront jellemzi. A második lépésben egy ún. aktiváló függvénnyel kezelhetőbb alakba hozza az így kapott számösszeget. Erre régebben a signoid[83] függvény volt a leggyakoribb (mely a 0 és 1 közötti tartományba korlátozta a neuron kimenetét), vagy a tanh függvényt használták, (ez -1 és 1 közé teszi a kimenetet, így szimmetrikusabb értéket kapunk). Ma azonban már a jóval egyszerűbb ReLU[84] az elterjedtebb aktivációs függvény az előnyei[85] miatt. Az aktivációs függvény által áll tehát elő egy neuron kimenete, ez lesz a következő réteg minden neuronjának elküldve. Tehát a hálózat csupán számok és számítások rendszere.

Tanításkor meghatározzuk egy-egy bemeneti adategyütteshez (mintához) az elvárt kimeneti értékeket, és mérjük, hogy ehhez képest milyen számokat produkál a modell. A tanítási folyamatban az elvárt értékek és a kapott értékek közötti eltérés minimalizálására törekszünk. Vagyis a mintaadatok az állandók, és azt számolja ki a gép, hogy a függvény változóinak (a súlyok és a biasok) milyen érteket kell felvenniük ahhoz, hogy a lehető legkisebb legyen az eltérés az elvárt és a kapott kimenetek között. Matematikailag ez egy függvény (ún. költségfüggvény) minimumának megkeresését jelenti. Ehhez a próbálgatós módszerhez egyszerű műveleteket kell elvégezni, de már kis modellek esetében is óriási mennyiségben. Ez okozza azt a nagy számításigényt, amely sokáig az ilyen rendszerek fejlődésének gátja volt. A fent ismertetett

[82] Erre azért van szükség, hogy úgymond el tudjuk tolni a kapott függvényt a koordináta-rendszerben, ha szükséges. Ugyanakkor ezzel a „bias" szóval utalnak sokszor az MI előítéletességére (torzítására, ld. IV.2.).

[83] Az ismertetett függvények képletét és koordináta-rendszerben való ábrázolásukat egy bővebb változat tartalmazza; úgy ítéltem meg, hogy ezek ismerete a továbbiakhoz nem szükséges, csak sokféleségük ismerete.

[84] Rectified Linear Unit, melynél a bemenet értéke változatlan, ha a kimenet pozitív, viszont 0, ha negatív lenne.

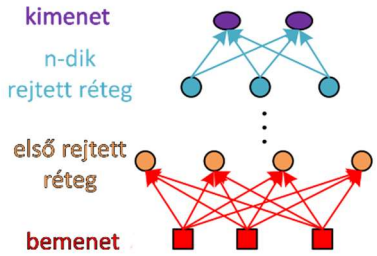
[85] A ReLU indoka matematikai, ezért nem részletezem. (Előnye, hogy nem okoz telítettségi problémákat, mint az előzőek, és a gradiens problémák feloldásában is hasznos; hátránya, hogy az értéke tetszőlegesen nagy lehet, tehát ezt a problémát külön kezelni szükséges. Erre a korábbi függvények mellett nem volt szükség.)

költségfüggvény-minimalizálás azonban egy mátrixművelet, melyben a részszámítások jó része független egymástól, tehát elvégezhető párhuzamosan. Ezért hozott az MI-ben áttörést a párhuzamosan végrehajtható egyszerű számításhoz tervezett processzorok (GPU[86]) megjelenése és terjedése – bár ezeket akkor még nem ehhez, hanem a számítógépes játékokhoz és a CGI[87] készítéséhez fejlesztették (mára már az MI-hez is fejlesztenek, ld. II.3.1.).

Az, hogy a rendszer csak számadatokat tartalmaz, egy „feketedoboz-jelleget”[88] ad a modellnek. Ez azt jelenti, hogy a belső rétegek tartalmát hiába is kérdeznénk le, sem a neuronok konkrét tartalma, sem a súlyok értéke nem ad értelmezhető információt úgy, mint pl. egy adatbázis-cella. A rendszer csak egészében használható, részei értelmezhetetlenek. Hasonlít ez ahhoz, amikor jegyzettömb (notepad.exe) programmal nézünk bele egy futtatható (.exe) fájlba: ebből nem tudunk következtetni semmire, pedig lefut a program. Ezért nem kérdezzük le a neuronok értékét: számunkra a belső rétegek „rejtettek”, így nevezi ezt a jelenséget a szakirodalom.

II.2.2. Más cél, más ML

Túlzottan nem érdemes belebonyolódni, de néhány eltérő modell vázolása elkerülhetetlen. Pillantsunk vissza a 2. ábrán bemutatott mély neuronháló (ANN/DNN) modellre, mely egyszerűsítve látható a 4. ábrán[89] annak érdekében, hogy a következő két modell könnyebben összevethető legyen. Látható is rajta, hogy ezen alapkonstrukció szerint modellekben az információ csak előre halad (előrecsatolásos 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.



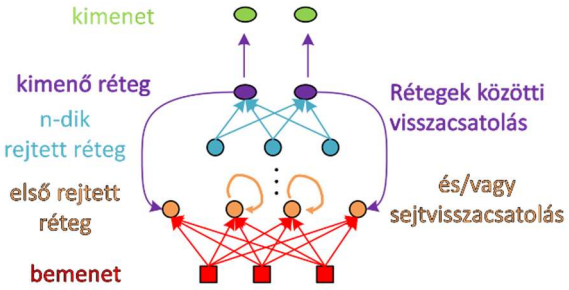
4. ábra: Az ANN-modell (saját átszerkesztés)

[86] Graphics Processing Unit (grafikai számítóegység). Míg a CPU-ban nagy számítási képességű magokból kevesebb van, a GPU-ban óriási számú egyszerű processzormag párhuzamos működtetésén van a hangsúly.

[87] Computer-Generated Imagery (számítógéppel generált képalkotás), de a 3D video animáció is ide tartozik.

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

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



5. ábra: Az RNN-modell (saját átszerkesztés)

A visszacsatolásos[90] 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 hosszú vektorú rövid távú memóriarendszer, ami az emberi rövidtávú memóriához hasonló dolgot hivatott megvalósítani. Hivatalos neve valójában inkább szójáték: „hosszú rövid távú 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 (LVSTM) kifejezés.

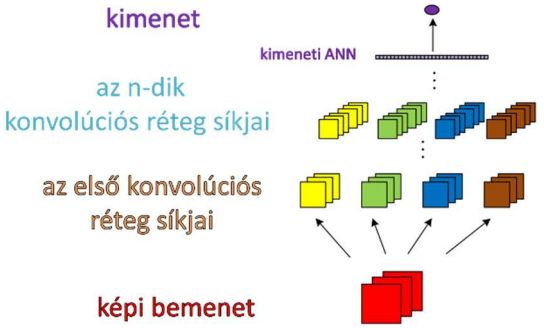
Az LSTM továbbfejlesztései képesek irányított módon megőrizni, illetve elfelejteni az információkat, ehhez memóriacellákat és/vagy kapukat használnak. Így ezek még hosszabb szekvenciák értelmezésére válnak alkalmassá. Az ilyen és a „kapuzott ismétlődő egységek” (*Gated Recurrent Units,* GRU)[91], illetve 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 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 inputrétegben is párhuzamos 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 tehát legfőbb különbség az alap ANN-modellhez képest. A párhuzamosság azonban

[90] Néhol „ismétléses neuronháló”, mely véleményem szerint nem egy találó kifejezés.
[91] Az ilyen rendszerekről alaposabb képet ad: [73].

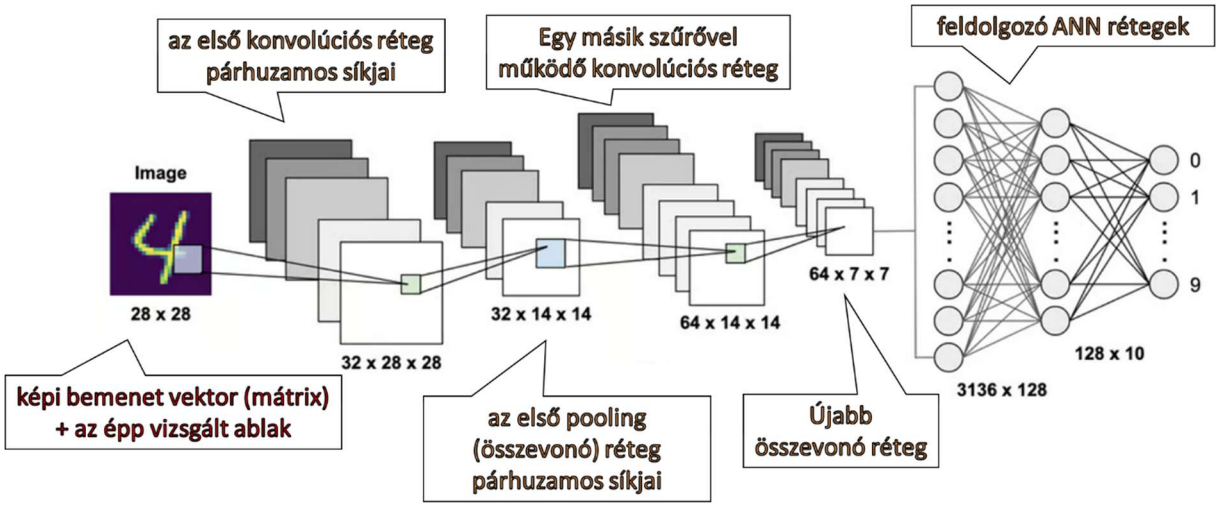
kevés a módszer megértéséhez, bele kell pillantanunk a működésbe. Ha ez itt nem sikerült, közérthető nyelvezetű forrásokat is adok meg.

A képet egy konvolúció nevű átalakítással tesszük a gép számára értelmezhetővé, ezért nevezik ezt a modellt konvolúciós neuronhálónak (*Convolution Neural Network*, CNN). A konvolúció matematikai művelete két függvény alapján egy harmadik függvényt állít elő. Ez az eredményfüggvény azt fejezi ki, hogy hogyan módosítja az egyik függvény alakját a másik függvény. Esetünkben a képmátrix képrészleteinek sorozatát egy ún. szűrőfüggvénnyel veti össze a konvolúció. A 7. ábrán [74] látható a képrészletet mutató „ablak", melyet kis lépésekkel, akár pixelenként arrébb tolva, végigmozgatnak a kép minden részén.[92] 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 módosulásai 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]



6. ábra: a CNN-modell egyszerűsítése (saját átszerkesztés)

Ám a konvolúciós képfeldolgozás csak több szűrőfüggvény lefuttatásával ad megfelelő eredményt, ezek különböző rétegekben működnek. Ezek között a rétegek között azonban érdemes



7. ábra: A CNN-modell alaposabb bemutatása (saját átszerkesztés)

[92] Ezért *csúsztatható ablakok technikája*ként is utalnak erre a módszerre.[75]

volt segédrétegeket is létrehozni a folyamat gyorsítása érdekében. Ezek a köztes, összesítő ún. 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, pp. 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 fontosságú 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 néhány újítása

Mivel néhány havonta jelennek meg új fejlesztési lépésekről publikációk, ezekből néhány bevált modellt ismeretetek. Felhasználói társadalmunkban egyre inkább népbetegség a figyelemzavar, a gépek koncentrációs képessége viszont egyre jobb. E téren a nagy áttörést az ún. 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 koncentrál.[80] Így egy adott információhalmazt több szempontból lehet vele elemeztetni, különböző kérdések mentén. Korábban egy LSTM-RNN-modell könnyen belezavarodott például egy hosszabb mondat helyes fordításába, mivel a gép a közeli hasonlóságokat és az ismétlődéseket nehezen kezelte. A megoldás elég új [81], és a módszer természetesen itt is a súlyozásokhoz kapcsolódik. Úgy oldották meg, hogy nem csak kulcsszavakhoz rendeltek magasabb súlyt, hanem szókapcsolatokhoz, vagy más szövegegységekhez is. [93] 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 figyelemmechanizmusnak ismertebbé vált az egyik fejlesztése, mint ő

[93] 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.)

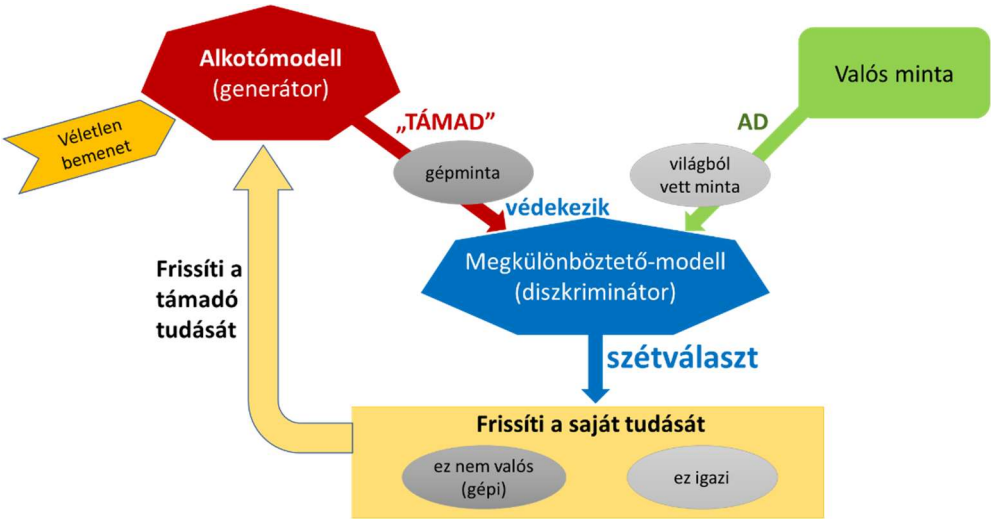
maga. Ez a Transformers rendszer, melynek névadója ugyan a Google volt, akinek a fejlesztői 2017-ben publikálták ezt a modellt, viszont a modell neve a Microsoft által vált közismertté, hiszen erre utal a GPT rövidítés a közismert ChatGPT (ld. II.3.4.) szolgáltatásban (*Generative Pre-train Transformer*). A Transformers modell azt oldotta meg, hogy sokszor nem ad megfelelő eredményt, ha a figyelemsúlyokat csupán átlagoljuk. Az új modell ennek elkerülésére többféle figyelemösszesítés („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.[84] Az eredményt tovább javította, amikor a többfejű figyelem mellett a modellt 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 (*Generative Adversarial Network,* GAN),[94] 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 alkotnak egy szép képet, vagy például hexameterben, vagy adott rímképlettel írnak verset. A névben az *adversarial* szó arra utal, hogy egy versengő modellről van szó, ami azt jelenti, hogy a GAN-ban két MI „ellenségként" viselkedik egymással.

[94] Az angol kifejezés szó szerinti „ellenséges" fordítását használva (mint több helyen), itt is kissé megtévesztő kifejezést kapnánk.

Az alkotómodell (generátor) a támadó, ez próbál olyan mintákat létrehozni, mellyel megtéveszti ellenfelét, amint azt az 8. ábra mutatja. (Az ábrát ihlette: [86].)

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



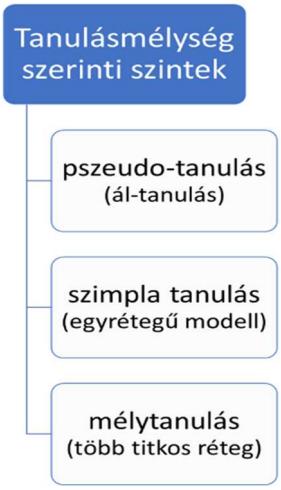
8. ábra: Egy generatív MI-modell, és annak „ellenséges” elemei (saját készítés)

nem egy, hanem két modellt szükséges finomhangolni, valamint a rendszer tanításához egyaránt szükségesek a valós és a hibás adatok (természetesen elkülönítve). Ezek rámutatnak arra a tényre, hogy az egyre komplexebb rendszerek tanítása is egyre bonyolultabb. Ez technológiavédelmi 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 pszihológiai és az információs műveletekben. Ám a jövőre nézve nagyobb jelentőségű, hogy hasonló versenyeztető módszer lehet alkalmas például kibervédelmi MI kiképzésére is.

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

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

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



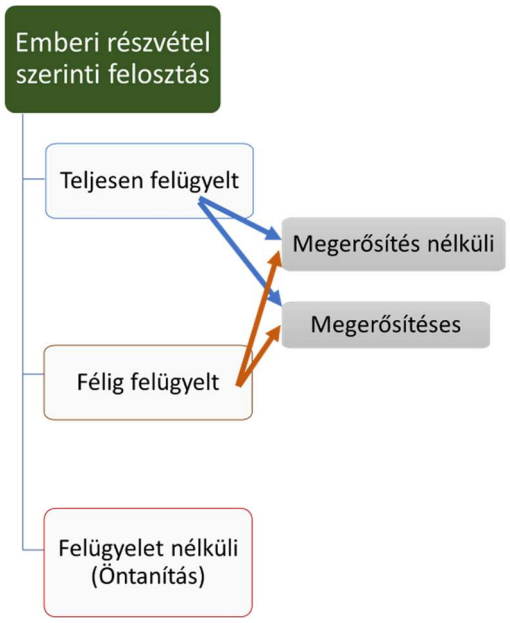
9. ábra: A gépi tanulás architektúra szerinti felosztása (saját készítés)

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 láthattuk, a mai 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:
   a) a DL több rejtett réteget használ;
   b) a szimpla DL csak egy belső réteget használ;
   c) egy réteget sem használ. Ezt hívom pszeudo-tanulási szintnek, de nevezhetnénk nulladik szintnek is. Abban az értelemben nem nevezhető tanulásnak, ahogyan a fentebb bemutatott modellek azok. Az ilyen hagyományosan programozott rendszer csupán a tanulás benyomását keltheti. Ezt fontossága miatt alaposabban majd a másik két főfogalommal összegzésben fejtem ki (V.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).[95]

[95] A gépi tanítási technikák számos további módon osztályozhatóak (ld. pl. a [87] sz. forrás 2. számú ábrája).

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.
- **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**: a tanulás közben nincs szüksége emberi beavatkozásra, a tervezése során alakítják úgy ki a rendszert, hogy erre képes legyen. Tehát nélkül fedez fel mintákat az adatkészletekben, és pontos eredményeket ad. A klaszterezés a felügyelet nélküli tanulás leggyakoribb alkalmazása.



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

- **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

Megnyugtatásul 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.2.3.) 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 a filmek vagy novellák öntudatra ébredő rendszereitől a mai technológia még fényévekre van.

Remélhetőleg a fenti néhány fontosabb technológia vázolása 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 jobbára együtt alkalmazzák (sok további fejlesztést is beépítve). Ám még ezek a hihetetlenül komplex és költséges óriásMI-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 szorítkozom, 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-szoftverek futtatására). Azok a fejlettebb CPU termékek sem nyújtanak megoldást, melyek egyre több magot tartalmaznak: azért nem optimálisak, 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 közvetkezők.

- Adaptált hardverek:
    - **GPU** (*Graphical Processing Unit*).[96] A más célra tervezett, rendelkezésre álló hardverek közül a kép- és videófeldolgozáshoz már régóta használt hardvermegközelítés, 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

[96] 1999-től jelentek meg ilyen grafikai célprocesszorok. Elsősorban játékokhoz és filmanimációs felhasználások elvárásai alapján fejlődtek, de a kriptovaluta bányászatból származó kereslet miatt is lökést kaptak.

használó rendszerek jól teljesítenek, a nagy gyártók azonban mégis kínálnak több más megoldást is.

- **FPGA** (*Field-Programmable Gate Array*). A tükörfordítás helyett hívhatjuk inkább a *felhasználáskor programozható logikaikapu-mátrix*nak, és még az 1980-as évekre nyúlik vissza (eredetileg semmi köze az MI-hez). Az egyszerűbb nevén szoftverprocesszornak 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 energiafogyasztású [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.

- Általános megközelítések:
  - **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 memóriahozzáférést, illetve az MI-sejtekben elegendő egyszerűsített számítást használnak.[91] Megjegyzendő, hogy bár a GPU terén még az amerikai szilíciumvölgy (az NVIDIA) a világ vezető cége, ám védelmi szempontból kiemelendő, hogy az NPU terén a kínai Huawei is az élvonalban van. Már 2017-ben felzárkózott az iPhone-hoz azzal, hogy a Kirin 980 mobiltelefon-processzor NPU-t is tartalmazott [92] (ez még inkább képfeldolgozásra készült), de szervert is építettek külön NPU-modullal (mely négy általános NPU-t intergrál).[93]
  - **Vegyes rendszerek**. Sokszor érdemes az MI-rendszerek mögé tervezett hardvereknél is vegyes megoldásokat alkalmazni, melyekben a fentebb felsorolt célhardverek 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 többmagos CPU-t egészítenek ki benne GPU és NPU elemek.
  - **RDU** (*Reconfigurable Dataflow Unit*), azaz egy „Újrakonfigurálható AdatfolyamEgység” technológiája még kevéssé ismert: itt a GPU-nál rugalmasabb alapokra próbálják helyezni a párhuzamosítást.[94] Vagyis a SambaNova cég ezen megközelítése egy MI-re jobban optimalizált GPU-nak is tekinthető.

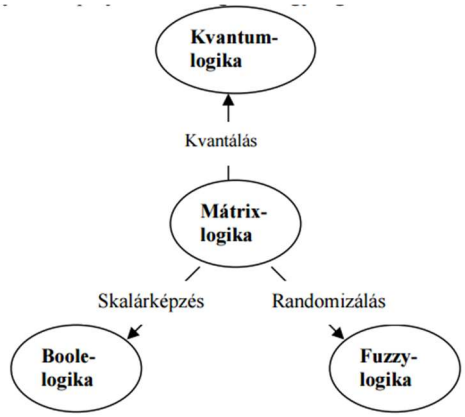
- Célhardverek, speciális MI-hez:
    - **Szoftverhez tervezett hardver**: egy további logikus architktúrális megközelítés, hogy egy adott MI-keretrendszer vagy -metódus számára optimalizálják a processzort. Erre példa a Loihi-2-es chipje (2021-től). Ezt kifejezetten a nyílt forráskódú LAVA nevű nyelvi keretrendszerrel való szoros együttműködésre tervezték, és a hivatalos tesztek szerint jóval hatékonyabbak a hagyományos processzoroknál.[95] 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.

Tekintsünk most 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,[97] szobahőmérsékleten, egyszerű elektromos impulzusokkal újrakonfigurálható alkatrészek hozhatók létre. Vagyis egy olyan anyagot alkottak, amelyből az elkészített processzorcellák igény esetén képesek négy féle 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

[97] A protonnal adalékolt perovszkit neodímium-nikelát ($NdNiO_3$) [96].

az FPGA-hoz hasonló, de annál sokkal nagyobb távlatokat nyitó megoldást fognak jelenteni. Ezen azonban jócskán túl is mutatnak. Mivel a cellák 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 mátrixlogika kapcsolata más logikákkal (saját átszerkesztés)

- **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 mátrixlogikát kidolgozó August Stern már 1970-ben felvetette a kvantumszámítógépek lehetőségét, jóval korábban, mint David Deutch fizikus, aki aztán a kvantumelméleti modell tovább-gondolásával rakta le a technológia alapjait.[98] Amint a 11. ábra [97] mutatja, a mátrixlogika kapcsolódásai már továbbvezetnek minket a következő témához. Visszatérve: a közeljövő technológiáját láthatjuk, egyelőre a kvantumszámítógépek alaplehetőségei is most bontakoznak ki, így még nem áll előttünk ütőképes, teljesen kvantuminformatikai 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 quibitekre[99] 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, életszerűbbekké válhatnának. Ez a tulajdonságuk, – karöltve azzal, hogy a kvantumgépek eleve nagyságrendekkel gyorsabbak, mint a hagyomá-

[98] Neki tulajdonítják kvantumszámítógép ötletét [97].
[99] Ez a kvantuminformatika alapegysége, mely a hagyományos bitekhez képest nem csupán a 0 vagy az 1 éréket veheti fel, hanem a kettő között bármilyen értéket.

nyos gépek, – egy valóban jelentős diszrupciót hozhatna létre. Érthető a vágy ennek birtoklására, a Google még nyílt forráskódú, hozzáférhető platformot is készített,[100] amely lehetőséget ad olyan MI-modellek fejlesztésére, melyek a kvantuminformatika egyedi lehetőségeit használják ki. A kutatás jelen állása szerint már számos kvantumtanulási algoritmus működőképes [98], bár leginkább a jelenlegi felhasználásnak igazán a hibrid modellek felelnek meg. Jelenleg a gyakorlati alkalmazhatóság felől az az irány tűnik legreálisabbnak, 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 széles körben 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 alapalkatrészeiből 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 iparággá 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 memóriaképességét igyekeznek adatok tárolására felhasználni, vagyis élő sejttenyészetekben 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 a jelenlegi fejlesztések (bár a biológiai alapú gépek környezetvédelmi aspektusa talán súlyosabb, mint az áramfejlesztés dilemmái). Ennek ellenére ezek használhatósága véleményen szerint egy igen távoli jövőbe tehető, ezért itt csupán említés szintjén elég volt utalni rájuk.
- **Biológiai struktúrákba ágyazott informatika**. Ez a kutatási irány az előzőre csak nevében hasonlít, de attól alapvetően tér el. Itt nem kiragadott alapegységekből építkeznek, hanem működő élőlényeket egészítenek ki (bővítenek) informatikai berendezések beléjük ültetésével. Ezeket jelen kutatás nem taglalja, – bár a jelen hipotézisek vonatkozásában is számos felvetést implikálna, ám – ezek tudományos vizsgálata megduplázná a terjedelmet. Néhány

[100] https://quantumai.google/

fontosabb kutatási terület alábbi említése is elegendő az ebben rejlő potenciálok érzékeltetésére, ugyanakkor a túlzott elvárások letörésére. Megjegyzendő, hogy a kifejezésben néha az elektronika szót használják, de pl. egy szívritmusszabályzót én nem sorolnék az itt releváns területhez; nincs kialakult elnevezése (lehetne pl. *Informatics Embedded in Biological Structures*, IEBS).

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 művégtagok erősebbek lehetnek az eredetinél, a műérzékszervek érzékelési tartománya tágabb lehet a biológiai érzékszervekénél, 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 szóltak, 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 szuperokos gép-ember hibrid ugyanis birtokolná az emberi alany születéskor kapott képességeit is, tehát agya által rendelkezne erkölcsi, kreatív, művészi képességekkel és valós érzelmekkel is, ezáltal egyből meg lennének oldva az erősen emberi tulajdonságok (emberségesség) gépiesíthetőségének nehézségei is. Természetesen ezzel együtt az ilyen számos aggályt felvetne, melyekre most nem térhetek ki. Sokan az evolúció

következő lépcsőfokának tartják az így létrejövő szimbiózist, melyre inkább a „szingularitás” (ld. I.3.3.) kifejezés terjedt el. A saját kutatásaim a témában részlegesek, 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.

Az MI fogalma szempontjából megállapítható, hogy nem várható, hogy hatást gyakorolnak rá a jelen hardverei, sőt a jövő felvázolt megoldásai közül is csupán a biológiai stukrtúrákba ágyazott informatika, 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 egy külön megnevezésnél, lehet pl. a mesterségesen kiterjesztett emberi kogníció (*Artificially Extended Human Cognition*, AXHC).

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

A fuzzy logikában rejlő szemlélet jelentősége az informatika alapjait érinti, hiszen a világ egy, a korábbinál használhatóbb leképezéséhez ad matematikai alapot. Itteni ismertetését ez a sajátos leképezési mód indokolja, hiszen emiatt 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 leegyszerűsítéstípusban, 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 összakapcsolása „kizáró vagy” (xor)[101] művelettel legtöbbször túlegyszerűsíti a valóságot.

Az imént 1. számmal jelölt problémát a 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

[101] 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”.

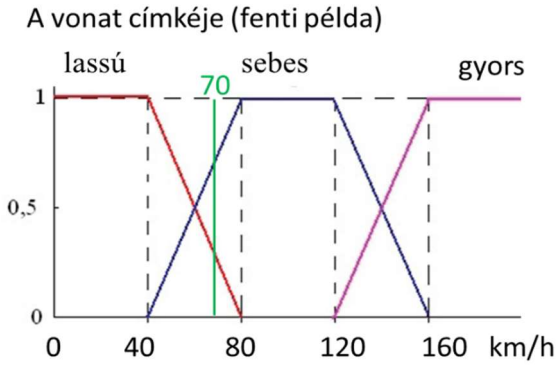
tartományt szeretnénk leírni, akkor a nyelvi változók a *világosszürke* és a *sötétszürke* lesznek. Egy közepesen szürke árnyalat tartozhat mindkettőhöz, – csakhogy ezen a ponton a példa nem a legmegfelelőbb. Jobb példa, ha a *sebesség* a nyelvi változó, és annak címkéi lehetne a *lassú*, a *sebes* és 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, példánknál maradva 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 alapmódszer tehát az adatokat nem éles adatként dolgozza fel, hanem

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

A 12. ábra[102] mutatja a *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 (≤ vagy ≥) 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 0 és 1 közötti értéket, amint látható az egyik címke (*sebes*) esetében növekvő, a másik (pl. a *lassú*)



12. ábra: A sebesség fuzzyfikálása példán (saját átszerkesztés)

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

esetében csökkenő mértékben. Így egy 70 km/h-val mozgó jármú „inkább sebes, mint lassú" értéke a két függvény együttes figyelembevételével[103] kezelhető. Így válnak számolhatóvá bizonyos nyelvileg amorf fogalmak, mint a „kicsi-nagy", „erős-gyenge" stb. Ez azonban nem oldja meg a leképezés mérhetőségi problémáit, vagyis az „erkölcsös" és az „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 lináris 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`[104] (ahol a *kicsi* és a *nagy* fuzzy halmazok értékei).[105] Vagy egy sebességszabályozást végző algoritmust, mely a fenti példa halmazait használja: `HA[[fékhőmérséklet = meleg] ÉS [sebesség = NEM villámgyors]] AKKOR [féknyomás == kissé csökkentve]`.[106] Ez a példaalgoritmus 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 számadatokra, melyekkel tovább számolhat a 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 a kiszámítása.[107] A fenti programpéldában erre utal a „*kissé csökkentve*" címke, mely tehát nem csupán egy „*csökkenteni* XOR *nem-csökkenteni*" választás, hanem egyből egy automatikus csökkentésmennyiséget 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.3.2.), akkortájt, amikor a kognitív tudományok és az MI fogalmait is meghatározták. Mint láthattuk, a fuzzy modell újfajta hidat

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

[104] A kódszerű részeket szándékosan fix szélességű karakterrel szedtem érzékeltetve, hogy ez nem rendes szöveg.

[105] A módszer megalkotójának példája. Ld. [106, pp. 29].

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

[107] Több ilyen módszer algoritmikus leírása (a fentebb említett színpéldán is): [107].

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 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).[108] A fuzzy technológia kirobbanó sikerére Japánban került sor 1987-ben, amikor a Hitachi elkezdte használni egy vonat vezérléséhez, az Omron pedig kereskedelmi forgalomba helyezte a világ első fuzzy alapú szabályozóját, sőt már 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 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) fontos szerepet játszik, hiszen a

[108] A következő történeti adatok forrása: [109].

zajcsökkentéstől kezdve a pontatlan egyezések kereséséig sok területen bevethető. A döntéstámogató rendszereknél az emberekéhez közelebbi döntéseket javasol. Az adatbányászatban és a prediktív modellezésben is kiemelkedően funkcionál, hiszen kezelhetővé teszi a nagy adatkészletek bizonytalanságát és a csupán 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 eredmény születhet.

Csakhogy 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 a definíció újragondolásakor nem szükséges figyelembe venni. A fentiek alapján világossá válhatott, hogy a világ leképezésének szempontjából is mennyire jelentős ez a szemlélet, tehát a humánvirtualizáció ismertetésénél (IV.2.2.(1).) felhasználható.

II.3.3. A biológiai ihletettségű informatikától a rajintelligenciáig

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 agykutatás 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 ötleteket meríteni, a biológia számtalan felfedezését sikerült már felhasználnia a mérnököknek, sőt egy új tudomány, a bionika (*bionics*) kifejezetten az ilyen lehetőségek számbavételére koncentrál.[118] Mivel az élet működési módjának megoldásait sem a Véletlen, sem a Teremtő nem védte le szabadalmi joggal, ez egy beláthatatlan méretű ötletbankként 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: [109] (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 részecskeraj 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 a 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, így számos optimalizációs feladat megoldására kínálnak a korábbiaknál sokkal hatékonyabb megoldást. Az egyik legrégebbi alterületről van szó: már a kilencvenes évek elején megjelentek olyan algoritmusok, melyeket a hangyáknál felfedezett feromonalapú kommunikáció inspirált. A 2000-es évek elejétől a hangyaboly-optimalizációt számtalan területen használják,[110] még kibervédelemben [121] is. A biológiai közösségek működéséből számos olyan algoritmusötletet sikerült meríteni, melyeknél az MI is segítheti a számítógépes implementáció hatékonyságát. A terület áttekintéséhez egy listát állítottam össze, ám részletesebb kutatását elhalasztottam, itt csak az elkészült felsorolást közlöm:

- a méhraj-intelligencia (pl. a naperőművek optimalizálására); [122]
- mesterséges halrajalgoritmus; [123]
- a szentjánosbogarak (fénylegyek) párzási motivációját használó modell; [124]
- pl. a pillangóalgoritmus; [125]
- a kakukk tojásrakása a Kakukk-keresésben; [126, pp. 105–116][111]
- a farkasvadászat a Szürke Farkas Optimalizálóban; [127]
- érdekes továbbá ezek rendszerezése is [128], de számos egyéb ígéretes kutatás is folyik.

Ezek mellett a biológiai szimbiotikus együttműködések tudatos átültetéseit is fel lehet használni [129], ezek az együttélés (*mutualizmus*), a *kommenzalizmus*,[112] vagy a *parazitizmus*, sőt

[109] Az alábbi listában szereplő területekről egy rövid átfogó képet ad: [119].
[110] A könyv első verziója 2004-ben látott napvilágot, melyben a szerzők számos felhasználási területet is bemutatnak.[120]
[111] Az itt felsorolt több algoritmus mellett szerepel ez is ebben a műben.
[112] Populációk olyan kapcsolata, amely az egyik fél számára előnyös, a másiknak közömbös (pl. az ürüléktermelő állatok számára közömbösek az ürülék feldolgozó organizmusok).

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 [4], amelyre én Szinergikus MI-t javasoltam (I.2.1.).

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 „rajmodell".

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 entitások 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 entitásai 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 az egyik hangya élelmet talál, akkor feromon jelzései által a többiek akkor is képesek lesznek hozzájutni és a boly egészét táplálni vele. És a boly tovább él akkor is, ha a keresés vagy a megszerzés közben meghal pár száz munkás hangya. Ennek katonai adaptációjánál tehát a cél elérhető úgy is, ha néhány drónt kilőnek a drónrajból.

Informatikai megvalósítás esetén az állatokhoz képest sokkal tökéletesebben adható át információ. Ezért egy MI-raj entitásai egymástól sokkal többet tanulhatnak, mint amennyit a legtöbb élő raj egyedei képesek egymásnak tanítani. Vagyis egy mesterséges raj esetén ilyen közös tapasztalatokból a cél elérésének leghatékonyabb módja gyorsabban csiszolódhat ki. 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, ahogy a raj-MI entitásai utódaikban módosítják saját programkódjukat is, így optimálisabbá tehetik azt a közösség tapasztalatai alapján. E rövid összefoglaló végén megemlítendőek a hierarchikusan és nem hierarchikusan szervezett rajok. Vagy van falkavezér és többféle rajszerep (pl. munkás és katona hangya) vagy nincs.

A raj számítógépes implementációját legoptimálisabban egy elosztott rendszer képviseli. Hierarchikus modellnél van egy központi vezérlőelem, ennek szerepét könnyen átveszi a „rangidős", 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 életképesebbnek tűnik, és ezzel visszaérkeztünk a rajintelligencia jövőbeni jelentőségének vizsgálatához.

Véleményem szerint a komplex MI-rendszerek architektúrájában alapvető szerepe lesz ezeknek a fajta szerveződéseknek. Általános MI-fejlesztések esetében bizonyosan számíthatunk rajképességek integrációjára, de már jelenleg is folynak ilyen jellegű kutatások.[130] Elsősorban több mélytanuló gép rajba szervezése jelent majd nagy lépést. Ezeknél a fentebb említett GAN-rendszerekhez 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 kifejezés[113]) alkotó entitásrendszerekből álló, összetett MI-raj, más néven *mesterséges mélyentitás hálózat* (*Artifitial DeepLearning Network* – ADLN, saját kifejezés) komoly lépés lehet a biztonságosabban és emberibben működő gépek, modellek felé. Ennek megvalósításához minden adott, nincsenek olyan problémák, mint pl. a kiterjesztett ember koncepciónál, tehát egy belátható jövőn belüli megvalósulása borítékolható. Ezért is tartom fontosnak már most megkülönböztetni a raj- és a központosított modelleket a különböző MI-definíciókban.

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,[114] mivel pár éve hirtelen kiemelkedett az MI-szolgáltatások közül és igen divatos kérdéskörré 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 *információ*kezelés. Más szóval a technológia igazi ereje és újdonsága az, hogy most már szinte információ

[113] Ilyen értelemben saját, hiszen rákeresve olyan találatokat kapunk, ahol nem egy entitáshálózatot értenek alatta, hanem csupán az MI használatát az emberi közösségben. Pl. https://www.eitfood.eu/projects/eit-community-artificial-intelligence.
[114] A témáról részletesebben is értekeztünk: [131].

(= 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 kezdhetnek megvalósulni a „szimbiotikus MI" (nálam szinergikus MI, I.2.1.) 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 akár irodalmi, akár szleng stílusban fogalmazni, sőt még humora is van.[115]

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 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 munkafunkciók munkaerőpiaci szerepének lenullázódása is. A fejlesztéseket erősen motiválja az MIKT-rendszerek erősödése is, hiszen az azokban keletkező óriási mennyiségű szöveg feldolgozása kizárólag szövegértelmező gépekkel lehetséges. Az adattavak esetében ugyanis nem lehet lekérdezéskódot írni, mely a kívánt válaszra szűri a hagyományos (strukturált, relációs) adatbázist. A karakteralapú kereséseken túl szükség

[115] Még közelebb hozhatja az ilyen szolgáltatásokat minden felhasználóhoz a Microsoft Copilot funkció, mely a megszokott MS Office dokumentumokba integrál egy olyan párbeszédablakot, ahol természetes nyelven kérhetünk táblázatműveleteket, egy prezentáció legenerálását, vagy egy levél adott szempontok szerinti udvarias megválaszolását és hasonló hasznos dolgokat. ld. [132].

volt egy jelentésre alapuló, „intelligens” keresésre is, mely képes záros időn belül a megfelelő információt kinyerni egy adattó strukturálatlan halmazából.[133] 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. Például egy frappáns kritika szerint „*a nagy nyelvi modellek csak abban jók, hogy megmondják, hogyan kellene hangoznia a válasznak, ám ez különbözik attól, aminek a válasznak lennie kellene*” [20]. Ezért – bár látszólag írathatunk vele tanulmányt, összeállíthatunk prezentációt, alkottathatunk leírásból képet vagy videót stb., – az ilyen rendszerek teljesítménye ámulatba ejtő, de valójában csak egy nyelvi zsonglőrmutatvány, egy jól betanított szuper-papagáj tevékenysége. Egy híres író szerint amikor az emberek felismerik egy adott technológia korlátait, akkor a kezdeti varázslatos jelleg szétmállik[116] – bár ez most még arrébb van.

Az technológiához szükséges modellt a nyelvészet adta: a hagyományos, emberi szövegkezelés nyelvészeti szintjeinek megfelelően hozták létre az NLP rétegmodelljét.[134] Ezáltal karakterektől (vagy hangoktól) a szövegkörnyezet 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 karakterhatárig), 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).[135] Fejlettebb NLP-k képesek a szövegben rejlő hangulat érzékelésére, sőt, kicsit hibás mondat értelmezésére is.

Bár sokan vágynak arra, hogy az NLP egy mindentudó, szinte gondolatolvasó varázslógépként 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 kérdésfeltevési stratégiákat is érdemes terveznünk ahhoz, hogy összetettebb kérdésekre a géptől a megfelelő választ kapjuk. Mindez tehát kissé nehézkesebb, mint a hagyományos internetes keresőkérdések, ám ha jól csináltuk, és szerencsénk van, akkor az eredmény messze meghaladja a böngészős keresések hatásfokát. Ha nem jól csináltuk, vagy a gép nem tudja, akkor egyelőre rosszabb a hagyományos módszernél, és talán ez a fő oka, hogy az

[116] Arthur C. 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.

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 teszteléseim alapján[117] pontokba szednem a terjedés lassúságának okait, hiszen ez egyben rámutat a népszerűség várható visszaesésére is.

- A látványosan téves válaszok, valamint az olyan gépi hallucinációk, amikor úgy tesz a gép, mintha tudná a választ, ír valamit, ami azonban 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 a 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 ismeretének a hiánya 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 információműveleti 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 álhírrel való információs 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 motiválhatók, hogy az MI-hez szükséges paradigmaváltásra rászánják energiájukat és idejüket – nem érnek rá ilyesmire… (ld. vertikális tanulás IV.5.1.)

[117] Az itt leírtakat saját tesztelések, három saját workshop és egy saját cikk tudásanyagából állítottam össze.

Összegezve: később gyakran lesz szükséges a fentebb vázolt ismeretanyagra. 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.

II.3.1. Neurális adatbázisok

A mélytanulás megjelenésével kézenfekvővé vált egy olyan adatbáziskezelési rendszer fejlesztési igénye, amely mesterséges neurális hálózatokat használ az adatok tárolására, lekérésére és elemzésére. Ezt valósítják meg a neurális adatbázisok, amelyek nem táblákban tárolják az adatokat, mint a hagyományos relációs adatbázisok. Erre csupán érintőlegesen térek itt ki, mert a különféle rendszerek összemosódásának példájához (V.1.2.) kapcsolódik.

Ennél a megközelítésnél sokféle különböző struktúrában tárolható az adatokat, attól függően, hogy milyen típusú feladatokat kívánnak majd az adatokkal megoldani. Ez a megoldás részben a kevésbé ismert objektumorientált adatbázismodellt 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).[136] A használt neurális háló típusának kiválasztására nem érdemes kitérni, a fentebb említett számos modell megjelenik a szakirodalomban, de vektoros modellek vagy gráfalapú 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 anomáliaészlelési feladatokhoz mintafelismerő képességük révén, így egy adattó 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 hibakeresést és a magyarázatot, hiszen a pontos működése nehezen érthető.

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

Összegzés. Bár a kijelölt témák rövid összefoglalása nem kis kihívás volt, de sikerült a tervezett mennyiségi keret közelében tartani a K2, vagyis a tanulás és technológia kérdéskörének vizsgálatát. Nem csupán bemutattam azokat a modern technológiákat, melyek az MI igazi hatékonyságával szorosan összefüggenek, de ezek közös halmazát is elneveztem (ld. R2.1.). A mélytanulási modellek ilyen kivonatolt vázlata is bemutatta mindazt, amit tudni szükséges az alapvető félreértések elkerüléséhez, vagyis a P1 témakör helyes vizsgálatához. Végül olyan vonatkozásokat írtam le, melyek elsősorban a P2 védelmi témakörhöz kapcsolódnak. Kiemelem, hogy a fejezetben leírtakat eddig is sikeresen használtam az oktatás során, a bemutatás különböző mélységgel emelhető ki, így jól illeszkedik alapozó kurzusokba a katonai felsőoktatásban.

P1-gyel és P2-vel egyaránt kapcsolatos következtetések

R2.1: Érdemes egy gyűjtőfogalmat használni arra a technológiakörre, amellyel összefonódva képes az MI igazán jól teljesíteni (II.1.). Javaslatom az MIKT (Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák).

R2.2: A rajintelligenciának kiemelt jelentőséget kell adni mind az MI-fogalom, mind a védelmi szempontok vonatkozásában (II.3.3.).

P1-gyel kapcsolatos következtetések

R2.3 A „pszeudo-tanulás” fogalmának segítségével világosabbá lehet tenni a mélytanulás és a hagyományos (változókba, adatbázisba való) adatgyűjtés közötti különbséget.

R2.4. A szinergia megvalósítása fontos elvárás az MI-technológia felé és a definícióban megjelenítendő (tartalma: ember és gép egymást kiegészítve képes dolgozni, I.2.1.). Az erre való törekvés a természetes nyelvfeldolgozás kapcsán érzékelhető (II.3.4.).

P2-vel kapcsolatos következtetések

R2.5. A fejezet nagy része jól használható a katonai felsőoktatásban, jegyzetként vagy tananyagként.

III. AZ AUTONÓMIA ANATÓMIÁJA

Minden jó döntés alapja, hogy a döntéshozó megfelelően tudjon élni a szabadságával, ne pedig visszaéljen vele. Az intelligencia segíthet megítélni, hogy mi a helyes. Az intelligencia pedig a tanulás révén lesz képes erre az ítéletre. Ez az embereknél régóta így van, de a számítástechnika által a gépek is elkezdtek az ember helyett dönteni, újabban pedig tanulási képességük által képesek meg is ítélni bizonyos dolgokat, ami a szabadság érzetét kelti. Ezért az intelligencia és a tanulás elemzése után – ezek következményeként – szükséges megvizsgálni az autonómia fogalom tartalmát is, részben az MI-fogalomhoz vezető áttekintés (C1) részeként, de még inkább a védelmi (C2) cél alapján. Ezen két irány mentén redukáltam ezt a meglehetősen nehéz területet a K3 kérdéskör[118] rövid áttekintésévé. Az alapkérdés az emberi és gépi autonómia hasonlóságaira és különbségeire irányul, mely módszertanilag úgy ragadható meg, ha fokozatokban definiáljuk az autonómia megvalósulását mindkét esetre, majd összevetjük ezeket a szinteket. Ennek érdekében a vizsgálat során az emberi autonómia áttekintése (III.1.) és a saját szintek meghatározása (IV.2.) után tekintem át a gépi autonómiát, annak szokásos szintjeit (IV.3.), majd javaslok rá egy saját alternatívát (IV.4.). Ezek alapján tudom összehasonlítani a kétféle autonómiát (IV.5.). (Bár a témához kapcsolódhatna, de nem szeretnék a jogi vonatkozásokra kitérni.)

III.1. AUTONÓMIA AZ EMBEREKNÉL: 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 „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 megközelítésben[119] a következő meghatározás adható:

[118] K3: Az autonómia kérdésköre: *Hogyan ragadható meg az autonómia gépi és emberi megva-lósulása, mik a főbb jellemzőik és a főbb különbségeik?*

[119] Ehhez a kiindulási alap egy technikai (katonai) megközelítés volt, melyet teljesen átfogalmaztam.[3]

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 belefogalmazása 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. 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 vélem, érdemesnek tűnik egyből egy olyan felosztáshoz lépni, mely egy gépi autonómiával foglalkozó szerzőn[120] alapul. Az autonómia és a moralitás szintjeit itt csupán vázlatosan közlöm, viszont további szakirodalomi észrevételek[121] figyelembevételével, a forrásokétól eltérő sorrendben.

0. racionális autonómia (funkcionális moralitás): A döntéshozó mérlegel, és az általa legsúlyosabbnak tartott okok alapján cselekszik.
1. döntési autonómia (operatív moralitás): A döntéshozó nem csupán a meghatározott oksági mátrix alapján mérlegel és dönt, hanem döntéshozó képessége egyre optimálisabb döntéseket képes hozni (öntanulás által képes fejlődni).

[120] Az angol forrás a „döntéshozó" helyett kifejezetten az gépre utaló „ágens" szót használja.[137]

[121] Több kifejezést a magyar interpretátor terminológiája, vagy saját megközelítés szerint közlök.[30, pp. 137.]

2. erkölcsi autonómia (elvhű moralitás): A döntéshozó saját erkölcsi meggyőződése és elvei alapján választ a lehetőségek közül.
3. személyes autonómia (teljes moralitás): A cselekvő maga határozza meg magatartását a legbecsesebb értékei szerint, ami személyes értékrendet és életcélokat feltételez.

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

III.2. ÁLTALÁNOS NÉGY-TÍPUSOS AUTONÓMIAFELOSZTÁS (4TA)

Az iménti megfogalmazások a kérdéskör újragondolását igényelték, melyet a következő szempontok szerint végeztem el:

- Megközelítésemben szerepet kap a döntési vektor.[122] Ez az eredő vektor alapvetően a múlt tapasztalataiból (döntések következményeinek tanulásából), és a jövő felé mutató reményből tevődik össze (reméljük, hogy döntésünk a várt eredményt, vagy a legkisebb rosszat hozza). Ezek további részvektorokra oszthatók, ezeket vizsgálom, és a részvektorok eltéréseiből 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).
- 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 éles helyzeteket, ahol nem az számít, hogy az illető egy moralizáló beszélgetésben hogyan nyilatkozott. Ezt a terminológiában úgy jelzem, hogy viselkedéstípusokat társítok hozzájuk. Ezek természetesen morális típusok is egyben, ám ez a tettközéppontú szóhasználat a különböző autonómiatípusokatjobban pontosítja.

[122] Ehhez a Lewin-féle mezőelmélet adja az alapot.[138]

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

III.2.1. 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 érzékszervére, 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 megszokott eljárása szerint bezárni a részleget, nagyszámú leépítést okozva.

Az ilyen típusú döntés „belülről” nézve kicsit determinisztikus, esetünkben a „vagy bezárjuk a céget 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 sablonját használni, megoldást kezd keresni. Más esetben az érzelmi hátterű logikát sokszor az elme racionális oknak tünteti fel, ilyenkor nagyobb eséllyel negatív irányba mozdít. Például amikor a katona azért nem teljesíti a parancsot, mert gyávasága meggyőzi őt arról, hogy parancsmegtagadása a logikus döntés, és ebben a logikában egy hadbíróság kisebb rossznak tűnik („hátha csak lecsuknak, és túlélem”), mint az életveszély. (Pedig lehet, hogy józanabbul átgondolva épp a parancs mentené meg egyébként az életét, de ezt az adott helyzetbe szűkülve (egyszerűsítve) nem képes átlátni.)

III.2.2. 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. Ez a döntési típus is támogatható tehát szabályzások, eljárásrendek, tervek stb. által, de abban tér el az előzőtől, hogy egy összetett helyzetben sok várható (tervezett, szabályozott) és váratlan (egyedi) tényező együttes figyelembevételére is képes. Vagyis képes sajátos, és széleskörű tudást igénylő, valamint összefüggéseket figyelembe vevő döntésekre is. 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 „gyárrészlegbezárás" 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 helyzetről 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 elhanyagolására 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.

III.2.3. 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 a 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 organikusan 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.[123]

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 tévedésének esetleg súlyos következményeit), akkor pozitívan értékelhető. Egy tűzesetből vett példa erre másképp mutat rá: egy polgár kisegíti egy közepesen magas ablakon a gyermeket is és az időset is, nem pedig dönt a két eset között, hogy melyiket mentse meg. Így ugyan eltörhet a lábuk, de mindkettő megmenekül. Az illető nem ismeri az erre vonatkozó szabályokat, de az élet értékének törvényét kreatívan alkalmazva képes cselekedni.

Fontos, hogy azért szükséges itt egy valódibb kreativitás, mivel a jó döntést általában nem az adott dilemma kettősségének szintjén (vagy akár sok hasonló eset között) kell keresni, hanem képessé kell válni „egyet hátralépve" a problémát egy nagyobb térben újraértelmezni. Gyakran egy nehéz problémára 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 modellezhetőségének 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 választ kapunk attól függően, hogy az idősek tisztelete a helyi elvárt prioritás, vagy „a fiatalok előtt áll az élet" elve hatja át a közfelfogást. Ez alapján fogja valamelyiket előnyben részesíteni a mentésbe beszálló önkéntes, akire nem mondható, hogy rutin, vagy tapasztalat alapján dönt, csupán a közfelfogást alkalmazza.

III.2.4. 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 szélsőséges helyzetben jönnek felszínre, amikor nincs idő a döntésre, és kialakult rutin sem lehet rá. A kreativitás és racionalitás csak lenyomatként

[123] 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.)

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 „ő ugyan otthagyta 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 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[124] igazi áldozatot hozni. Érdekes, hogy sokan ezekben a hivatásokban (pl. katonai, rendészeti, katasztrófavédelmi, egészségügyi, pedagógiai) sokan lát-

[124] Azért használom ezt a szót, mivel nem pontosan a jaspersi értelemben vett „határhelyzetekre" gondolok, hanem sokkal egyszerűbb szituációkban is bekövetkező, váratlan, de felelős döntést igénylő helyzetekre.

nak nagy perspektívákat az MI számára. Itt megjegyzendő, hogy a hősiesség nem csupán kikerült az elvárásokból, hanem sokszor szabályellenes is, egyébként joggal. Például afenti tűzesetnél könnyen lehet a döntés hátterében 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.

III.3. A GÉPI AUTONÓMIA ALAPVIZSGÁLATA

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.

III.3.1. 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 irányítják (pl. összerezzenés hirtelen erős hanghatásra), de a gépek esetében másfelől kell közelíteni a fogalomhoz. Nézzük azonban először azt a megközelítést, hogy mi az alapvető különbség a kétféle 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.*

Az autonómia rövid meghatározását átvettem [12], az automata megfogalmazása saját, és a definíciók arra hivatottak, hogy a kettő egymás mellé állítása rávilágítson szétválasztásuk szokásos módjára. Tanulmányomban, – főleg ebben a fejezetben, – általában kénytelen vagyok az autonómia szót is használni, annak ellenére, hogy nem tartom pontos kifejezésnek. Ugyanis ez terjedt el, tehát segít a magyarázatban, és – az emberi autonómiával való összevetésben – emellett van különbség az „önmaguktól" működő gépek két típusa között, tehát jogos a hagyományos automatikától való megkülönböztetés valamilyen formája.

De az automatikát is fontos alaposabban elemezni a gépi autonómia szokásos megközelítésének továbbgondolásához. A technika fejlődése sokféleképpen osztható fejlődési irányokra és korszakokra, itt az egyszerűség kedvéért egy saját felosztást közlök. Ezek a generációk nem elsősorban az időben értelmezendőek, mivel sokszor egy-egy területen már a második generációs gépekről beszélhetünk, míg a másik területen éppen megjelennek az első megoldások.

0. A nulladik generációs gépek inkább csak ügyes eszközök, melyek évezredeken 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 technológia által kap energiát** működéséhez egy 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 vágtázó lovat lehagyni, 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 emberi értelemben soha nem mérlegel, minden döntést az ember előre meghoz. A gép nem szabadon választ a lehetőségek közül, hanem választásának pontos feltételei az előzetes emberi döntés részei, melyeket a gépi döntés csak érvényre juttat. 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[125] viszont itt alaposabban ki kell térnem, mivel ezek tűnhetnek önállónak. A többi automatához képest jelentős eltérés, hogy a programkód képes kezelni egy sor

[125] Programkód (script) alapján működő automaták.

működés közben bekövetkező eseményt is. Ezt háromféle emberi beavatkozással vagy valamilyen szenzoradattal lehet befolyásolni. A legtöbb elágazást (döntést) a programozó (1) határozza meg. Sokféle döntést az üzemeltető operátorok (2) állítanak be, a programozók által előkészített beállításhalmaz lehetőségein belül (pl. hogy a képernyő zárjon le 10 perc inaktivitás után). Bizonyos esetekben a felhasználót (3) kérdezi meg a rendszer, mielőtt továbblépne („biztos, hogy kilép mentés nélkül?"). Ezen felül szenzoradatok alapján is dönthet a gép pl. egy processzor túlmelegedését észlelve üzemszerűen állítja le a rendszert.

Tehát elvileg emberek által determinált az ilyen rendszer is, ám kiszámíthatatlansága nagyobb, mint a korábbi generációké vagy automatáké. 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 hardverszintű bonyolultsága vezethet váratlan esetekhez: a lényeg, hogy a kb. harmadik számítógép-generációtól (a tranzisztoros számítógépekről) kezdve megjelennek az emberi elmével átláthatatlan bonyolultságú gépek. Minden üzemeltető tudja, hogy „némely gépnek lelke van", de azt is tudják, hogy ez mindig valamiféle működési vagy kezelési anomália, mely sokszor ismeretlen marad.

III.3.2. Autonómiaszintek, melyek valójában automatikaszintek

Itt érkeztünk 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 chatbotok 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 kutatástechnikai probléma is felmerült bennem. Alább ezeknek járok a végére, mivel fontos tanulságokat rejtenek:

I. Miért nehéz általános megközelítéseket találni?
II. Hogyan lehet a különféle okosgépek 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 kezeltek minden ilyen kérdést.[139] 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 szükség 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*."[12]

A fenti (I-V.) kérdések egy részére választ adhat az egyik konkrét szintezés elemzése. Népszerűsége és ismertsége miatt célszerű az autók besorolásait venni példaként. A szintek ismertetése előtt hangsúlyoni kell, hogy ez a hivatalos J3016 szintezés [140] valójában a vezető támogatásának szintjeit határozza meg. Röviden [141]:

0. szint, a vezetésautomatizáció teljes hiánya. Lehetnek azonban különböző MI-alapú támogató, figyelmeztető funkciók (pl. holttérfigyelő, sávelhagyásra vagy elalvásra figyelmeztetés), sőt egy automata vészféktől is nulladik szinten marad a kocsi.
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 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 a 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 automatizálásáról 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 autonómiája divatszóvá válik, hiszen pl. a szabvány magyarázatainak nagy részében is autonómiaszinteknek hívják ezeket.[142] Sőt, a lista alapján felmerül, hogy az autonómia szó használata kissé korai, pl. igazából az 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 autonómiaszintjeihez.[143] 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 tolódik: a különböző felhasználási területek szerinti autonómiabesorolásokat 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 autonómiaszintezések. Ennek igazolására én a logisztika [144] és az egészségügyi robotok [145] 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.

III.4. GÉPI-AUTONÓMIASZINTEK (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.

III.4.1. A hétszintű felosztás

A gépi autonómia szintjeit ezért más irányból vizsgálom. Az autók példájánál maradva: az autó fedélzeti okosrendszere által vezérelt adaptív fényszóró tévedéséből is lehet egy MI által okozott baleset, ha elvakít egy szembejövőt (pedig ez a funkció nem emeli a járművet még a J3016 szabvány első autonómiaszintjébe sem, hiszen nem szól bele az irányításba). Ezért ahelyett, hogy ilyen eseteket egyenként beemelnénk a J3016 szabványba, inkább olyan megközelítést javaslok, melyben minden MI-funkció összefügg a biztonsággal. Ennek érdekében jó lenne, ha az autonómia szintje közvetlen kapcsolatban lenne 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 egyenletes és egyértelmű összefüggés, de alkalmas az autonómia szintezésére az állatok esetében is. A másik kiindulási pont, hogy a részrendszerek 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 reflexszerűen reagálni képesek a külső hatásokra, de mindig ugyanúgy. Az emberi rendszerekkel vagy fejlettebb állati autonómiával nem állítható párhuzamba, hiszen még a reflexeink is bizonyos mértékig kontrollálhatóvá válnak tanulással.

   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 pszeudoautonómia szintje a múltat naplózó programoknál.** Először a tanulásról szóló (II.2.4.) részben említettem a pszeudotanulást. Erre utalva beszélhetünk pszeudoautonómiáról is. A jelenség lényege, hogy klasszikus programok neuronháló helyett változókban vagy hagyományos adattáblákban tárolnak el „megtanult" adatokat. Ilyenkor úgy tűnik, mintha tanulnának, ha ezek alapján változtatják a reakcióikat is.[126] Mivel ez amolyan papagájtanulás csupán, ezért ezek semmiképpen nem nevezhetőek autonómiának, hiszen nem helyzetfelismerés alapján hozza meg a döntését (ld. III.3.1. definíciója), hanem csupán egy mért statisztikára ad egy kiszámítható reakciót. Azért kiszámítható, mivel a változót (vagy adatbázismezőt) lekérdezve bárki megmondja, mit fog dönteni a gép. Megemlítendő itt a véletlenszám-generálás, mely azzal teszi a világ leképezését valóságosabbá, hogy képes egy beállított bizonytalanságot adni a rendszernek.[127] A véletlen fogalma azonban a valóságban sem tekinthető szabadnak, mitológiai hasonlattal élve: Fortuna vak.

[126] Erre részletesebben visszatérek a pszeudo-MI fogalmánál (V.1.4.). Ott bemutatom a vásárlási szokások pszeudotanulását, ahol pusztán egy-egy adott címkével kapcsolatos kattintásokat mérve ad a rendszer egy-egy címkéhez valamilyen súlyt (ami idővel változik), és ez alapján javasol mindig más, releváns terméket a vásárlónak.

[127] Az előző lábjegyzet példájánál maradva: a termékkategória címkéjét a statisztika dönti el, az ajánlott termék ezen belül lesz kisorsolva.

Ez a szint a 4TA-lista „egyszerűsítő autonómia” szintjének alapeseteivel állítható analógiába. A pszeudoautonómia 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 neurális tanulás.** A rendszer géptanulás által támogatva képes sematikus intellektuális döntések meghozatalára: járművet parkol le, vagy egy gyártási folyamatot vezérel, de megjelenik ez a régebbi nyelvfeldolgozó szolgáltatásokban is (ezeknél a cselekvés a válaszadás), és sok más rendszerben. Együttműködő szolgáltatásoknál az optimalizációt az MI hatékonyabban támogatja, mint a pszeudo-MI-rendszerek, de abszolúte nem kreatívan.

Ez a gépiautonómia-szint is széles spektrumot takar. Egyszerű megvalósításai a 4TA-listában az emberi egyszerűsítő autonómia döntéseivel állíthatók analógiába, ám fejletteb megoldásokkal bíró béta rendszerek a 4TA összesítő autonómiájának alsó (szimplább) rétegeit is elérhetik. 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 biztonságot, mivel tervezéskor az egyediség határai a megtanult (emberek által megszabott) sémahatárokon belül maradnak.

Érdekes módon ezt a biztonságot a „lelki” térben is el lehet érni, nem csak a fizikaiban. Jó példa erre, amikor generatív rendszerek képesek pl. a sértő dolgokat elkerülni (ld. (IV.2.). Érdemes a kettőt összevetni, ezt három tényező szerint teszem meg. (1) A lelki tér kevésbé objektív, tehát bár a leképezés nehezebb, viszont jobban szimplifikálható. Pl. nem szükséges mindenki lelki világára tekintettel lenni. (2) Olcsóbb, hiszen csupán virtuális szolgáltatásként kell megvalósítani. (3) A tudásuk valójában elegendő, hiszen a kapott kreativitásuk korlátaiban jól teljesítenek.

A fizikai térben viszont ennek az autonómiaszintnek a fejlődése és terjedése jóval nehezebb, hiszen ott a fenti három tényezőnek épp az ellenkezője érvényesül. (1) A fizikai biztonság objektív. Minden ember testére vigyázni kell, és plusz a fizikai tér minden tárgyára is. (2) Az egyedeket, amelyekbe ilyet beépítenek, drágább legyártani. (3) 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” – ezzel szemben egy fejlett szoftveres fejmegjelenítés mimikája lenyűgözi. Ilyen gamma autonómiájú rendszerek a fizikai térben akár idősgondozó vagy ápoló robotokat is vezérelhetnek, de bonyolult vízi vagy légi manővereket is önállóan 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 a rendszer sokféle szempontot képes figyelembe venni döntéseinél, sőt képes lehet akár a következményeket szimulálva („a lehetőségeket átgondolva”) kiválasztani a legideálisabb cselekvési irányt. 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.1.5. és VI.2.4.), melyek ugyan hozhatnak az ember számára ártalmas döntéseket, akár emberéletek kioltásáról, ám ez nem különbözik bármely reguláris katona tűzparancsától. Vagyis épp úgy, mint egy közkatona, egy MI-modulokkal (főleg képfelismerés) működő katonai rendszer is a 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 az uradalmi 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. Úgy tűnhet, hogy döntenek, pedig csupán néhány statisztikai optimum eredőjét 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 sokmodulos, hierarchikusan összekapcsolt MI-struktúra. Ez azonban egy világosan lehatárolható döntési, illetve cselekvési tér, nem a valóság végtelen lehetőségeinek tere. Más szóval csakis zárt rendszerben, és csupán a delta szintű komplex feladat ellátása működik megfe-

lelően az ilyen autonómia. Ezzel a zártsággal a rendszerhibák miatti balesetek jobban elkerülhetőek, valamint a szabályozási nehézségek is megragadhatóbbak. Például bányák, kikötők, raktárak automatizálása kezdődött meg ilyen irányban, és ezeken a területeken a delta szintű autonómiák terjedése gyorsabban várható. A belátható jövőben megvalósulhatnak összetettebb delta szintű autonóm rendszerek is, pl. egy ember nélküli fuvarozás is: a felpakolástól a közúti eljuttatáson át a lepakolásig.

Ám biztonsági intézkedések ilyen környezetben is szükséges lesznek. Például, hogy vészhelyzeti jelszóval bárki felfüggeszthesse a működést, illetve a váratlan helyzeteket emberi operátor oldja meg. Zártak a tesztkörnyezetek is, ahol lehetőség van a gépesítés kihívásait elemezni, így az érzékeny területeken való MI-használat hosszú ideig csupán a delta szintű tesztkörnyezetben képzelhető el. Ilyen terület például az egészségügy, ahol robotfelcserek 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 ember felügyeletével gamma szintre visszaléptetni. Például egy városi busz- vagy taxirendszer üzemeltetésekor kell valaki, aki a szükséges esetekben felülbírálja a rendszer döntéseit. Ilyen megoldással próbálta például a Cruise taxivállalat biztonságosabbá tenni önvezető járműveit.[146]

Azonban még ezek a komplexebb megoldások sem lehetnek képesek a 4TA harmadik típusára, azaz a szabálymögötti autonómiára; ezek is csupán az összesítő autonómia fejlettebb szintjét jelentik. A gamma rendszer példáját továbbgondolva itt a rendszer a különféle beosztott dolgozók, valamint az őket koordináló „brigádvezető” dolgozó együttes autonómiájával rendelkezik. Hozhat döntéseket a brigádvezető, ám csak felettesei elvárásai mentén, esetleg képzettebb szakemberek ellenőrzése mellet. 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. Ezt alátámasztja, hogy az Ipar 4.0 mentén egyre komplexebbé váló rendszereket tekintve is úgy tűnik, hogy ilyen irányban haladnak az innovációk, melyek autonómiája a jelenleg ismert elvekkel működtetve csak a delta szintet érheti majd el.

5. **Epszilon 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 szabálymögötti 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 átlépnénk a hipotetikus szférába. Jó példa erre az az irodalmi fikció, ahol emberi jogot szerezhetett egy, az emberhez hasonló tudatú gép.[147]

Ám a rajintelligencia 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 szabálymögötti autonómiájához nagyon közeli szintet.

6. **Zéta szint: autonómia 2.0.** Mivel az intelligencia esetében felvetődhetett az ASI-fokozat, -ezért következetes az így elképzelt jövő autonómiával kapcsolatban is felvetni hasonlót. Az evolúciós szintlépésből létrejövő új szuperentitás magasabb autonómiát is vindikálhatna, melyet a mi korlátos elménk úgysem lesz képes értelmezni. Ezzel az emberfelettiség igen problémás nézeteihez érkezünk, ahol nem egy emberi faj, hanem egy új „faj” lenne emberfeletti. Az ilyen nézetek egyrészt a történelmi tapasztalat alapján nehezen legalizálhatóak: egy „fejlettebb szabadságértelmezés” mások nézőpontjából mindig 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 statisztikái és matematikai modellek által feloldhatóak lennének az ősi morális paradoxonok. Pontosabban csakis az emberiesség megszüntetésével lehetne ezeket a paradoxonokat feloldani. A fuzzy logika (II.3.2.) az ellentétek bizonyos mértékű keveredését ugyan kezelheti, de olyan szimbolikus (talán metafizikainak mondható) jelentést, hogy pl. a víz egyszerre pusztító és éltető princípium: ilyet nem képes digitálisan leképezni. Tehát hiába irtaná ki az emberiséget egy ilyen elszabadult szuperautonóm MI, valójában úgysem haladná meg az ember morális szintét (a pusztítással főleg nem). Ezáltal véleményem szerint egy emberfeletti „autonómia 2.0” lehetetlenség; sem gép, sem ember által nem megvalósítható, mégpedig az önellentmondása miatt. Megjegyzendő, hogy ez kihat arra is, hogy az ASI-modell vajon mennyiben értékes: véleményem szerint, ha „okosság 2.0” még lehetséges is lenne, de 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 bűvöletében 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 belefektetett munkaórát, a sok jót, amit megmaradása esetén tehetne, többre értékelné egy (vagy sok) ember életénél. Az önzés ugyanis mindig logikus és optimális – az önzetlenség viszont általában paradox. Ilyen hipotetikus epszilon vagy zéta autonómiájú rendszerben persze felülbírálhatná egy alfa típusú kód az ilyen gépi önzést, ám az már nem gépi autonóm döntés lenne.

III.4.2. Autonómia és felelősség

Pár szóban a felelősség kérdésének egy filozófiai megközelítését vázolom, amely szorosan kapcsolódik az autonómiához – újra hangsúlyozva, hogy nem jogászként közelítek a témához. Először, a következő részben az előítéletességet járom körbe. A felelősség – mint láttuk – igazából csak a cselekvéssel összefüggésben értelmezhető. A tetteknek 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.[148, pp. 86–104]

A problémakör nagyon messzire vezet, és sok tudományágba beletorkollik: azt, hogy minden tettünk kihat a világra, az metafizikai, teológiai vizsgálat tárgya lehet, míg azon esetek elemzése, hogy valami probléma fellépése esetén ki a felelős, az a jog területe. A felelősség megoszlásának elvei a filozófiába vezetnek, a felelősség visszahatását az egyénre a pszichológia boncolgatja, kollektív esetben pedig akár a szociológia tárgya is lehet. A leggyakoribb, 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 okozatként az illető megbűnhődik. Ennek az ok-okozatiságnak a megvalósítása fontos pillére a társadalomnak, mert szinte mindenkiben belső elvárás, hogy egy rossz tett visszahasson arra a személyre, aki felelős az okozott 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 többrétű az emberi felelősség: tervezők, tesztelők, kezelők, szerelők, vezetői beosztásból irányítók, biztonságellenőrök lehetnek figyelmetlenek, tévedhetnek, sőt a résztvevők több apró hibázása is összeadódhat, melyek egyenként nem okoznának bajt. A bonyolult automatikák

mögött is felsejlenek az emberek az iménti lista szerint. Ám ez kezd eltűnni a neuronhálók egyre bonyolultabb „feketedobozaiban”. Mit tehetünk?

Egyik megoldás lehetne, hogy a gépet tesszük felelőssé. Ez azonban nem hozhat megoldást, mivel a gép nem tud bűnhődni, illetve büntetése nem fog az érintett emberek számára hasonló megnyugvást hozni, mint a bűnös ember megbüntetése. Lehet a gépben szimulálni szenvedést vagy becsvágyat, melyet megsértünk a büntetéssel stb. Kérdés azonban, hogy ez hány generáció múlva fog hasonló igazságérzést kiváltani a gépek által hátrányosan érintett emberekben és hozzátartozóikban, 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”.

Egy kis kitérőben nézzük meg ezt. Amikor egy gyerek szeretne több cselekvési szabadságot (azaz autonómiát), akkor azt nézzük, hogy érti-e, hogy valójában következményekkel járó, felelős döntéseket szeretne hozni? Az éretlen elméket igyekezni kell elzárni a teljes szabadságtól (pl. nem hagyni elöl 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 kultúráinkat. Abból indulunk ki, hogy feltételezzük, hogy minden ember képes felelősen dönteni (kivéve, aki nem[128]), és nem tiltunk meg általános dolgokat, nem vonunk meg jogokat, míg a kiindulási feltételezés ellen nem vét valaki. Egy műszaki megfogalmazással élve a felnőtt emberek esetében fekete listát alkalmazunk: bizonyos dolgokat nem mondhat, nem tehet (pl. bizonyos helyekre nem mehet be). Diktatórikus rendszerek esetében a feketelista 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 előállítása szükséges, mennyiben vesz ebben részt pl. az ego vagy a lélek (esetleg lelkiismeret) vagy a tudat. Ezért felelős gépek helyett

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

gyerekként kezelve őket egy úgynevezett fehérlistát is kell alkalmaznunk a gépek tevékenységi tartományára. Vagyis elsősorban azt mondjuk meg, mit tehetnek. Lekorlátozzuk a cselekvési tartományt: ezt jeleníti meg a gamma szintű autonómiánál említett zárt rendszer is. Tehát a hagyományos számítástechnika feketelistás szabályozásán túl csak azok a neurális rendszerek kaphatnak konkrét „jogot" bizonyos szintű döntésre (fehérlistás cselekvésre), 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 a véletlen szerencsétlenséghez hasonló kategória felé tereljék, ahol ugyancsak nem lehet emberi felelőst megjelölni.

III.4.3. Speciális esetek

A fenti szinteket a tisztán mesterségesen megvalósított döntéshozás alapeseteire fogalmaztam meg, de különféle speciális modellek, hardverarchitektúrák vagy elvárások ezt természetesen módosítják. Itt két speciális esetre térek ki: élőlénybe kapcsolt MI-re és magas rendelkezésre állás elvárására.

Egy MI-képességekkel kiterjesztett ember esetében az alsóbb szinteken is megjelenhet a 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 támogatná a hordozóját, vagyis csak szűk döntési terekben működhetne jól. Például az ipari vagy üzleti döntésekben segíthetne, ám ezeken kívül inkább összezavarná használóját, mintsem segítené a hétköznapi döntéseit. Vagyis véleményem szerint ki kellene kapcsolni a munkahelyen kívül az ilyen eszközt, hogy ne őrüljön bele a viselője.

A másik megvizsgálandó helyzet az úgynevezett „magas rendelkezésre állású" fokozatba sorolt ipari rendszerek esete (*High Availability System,* HAS)[149]. Ilyen elvárás mellett ugyanis kiemelkedő kérdés, hogy milyen szintű gépi autonómia mellett garantálható ez a besorolás. Ilyen rendszerek túlnyomó része védelmi szempontból is kiemelkedő, hiszen egy ilyen

rendszer jóval magasabb költségeit az indokolja, hogy leállása vagy sérülése komoly következményekkel jár – a komoly következmény pedig általában kihat az adott terület vagy ország polgáraira is (legyen szó erőműről, vegyi üzemről, bankról stb.). A témára még visszatérek (IV.2.&3., VI.2.4.)

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.

III.5. EMBERI ÉS GÉPI AUTONÓMIA ÖSSZEVETÉSE

Ezzel elérkeztünk az emberi és gépi autonómia összevetéséhez. Azokat a különbségeket emelem ki itt, melyek a C2 célhoz 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 embereknél a szabad döntés hiányával kapcsolatban szokott felmerülni, a gépeknél ezzel szemben épp fordított a probléma. Az emberben van egy belső szabadságvágy, amelyet normál esetben szeretne cselekedetben megvalósítani. 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.

   A gépek esetében azonban nem az okozza a gondot, hogy korlátozzák a szabadságában. Szimulálhatunk benne szabadságvágyat, de ez csupán egy hozzáadott modult jelent, nem pedig létéből következik, ahogyan az ember esetében. A gép visszajelezhet olyat, hogy korlátai miatt nem tudja megoldani az adott feladatot, de ez nem szabadságvágy. Az ember még

alapvetően fél a gép szabadságától, mint ismeretlentől. Pedig a fejlett térségekben a gépek eddig is hozzájárultak ahhoz, hogy több szabadideje legyen, könnyebben elvégezze a munkáját: tehát a gépek támogatták az ember szabadságvágyát. Az MI által még több szabadideje, könnyebb munkája, tehát több szabadsága lehet az embernek. Emellett helyes működés esetén a gép nem sérti senki 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, hogy az önérdek kizárása csak a gépek területén oldható meg, de embereknél nem. 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 szélsőhelyzetben az 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 izgalomtól a félelemig. Megjelenik ez az olyan helyzetekkel kapcsolatban is, amikor nem tudja, hogy a másik ember hogyan fog dönteni. Ám a másik ember döntése mögött képes tisztelni annak szabadságát, ezért az ilyen eredetű kiszámíthatatlanságot a legtöbb ember elfogadja. Megszokjuk, hogy az életet ilyen típusú bizonytalanságok sorának lássuk. 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, fel-fogni. Ez az ősi reflex úgy vélem, 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 Big Data felhőjébe vesző háttere). Ez a magasabb szintek esetében exponenciálisan erősödik, a félelmi tényezők hatásainak összeadódása miatt (ilyenek pl. a művészi alkotások, a konteo-elméletek, az információs műveletek stb.).
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 autonómia-szintek kapcsán már elemeztem. Egy fontos aspektust azonban érdemes megvizsgálni.

Elvi hasonlóság: A világ leképezésének helyességét nem tudjuk objektivizálni. Más szóval az ember által teremtett, szimulált virtuális szabadság hátterében lévő belső, leképezett világ éppúgy nem mérhető, mint az emberi szabadság lelki háttere. Erre nincs objektivizáló 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 objektivizáló eszközök pontatlanok, vagy a képletek nem vesznek elég tényezőt figyelembe. Itt elméletileg sem tudjuk a problémát megragadni, a leképezés szubjektív aspektusait nem lehet[129] objektíven leképezni. (ld. még IV.2.2.)

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. 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)?

III.6. Részösszefoglalás: A döntések és a hibázás szabadsága

Összegzés. A fejezetben a K3 kérdéskört csak a kutatás témáinak fényében volt hely megválaszolni. Kritikusan tekintettem az autonómiafelosztások jelenleg elterjedt verzióira, és saját felosztásokat határoztam meg. Az emberi autonómia felosztásakor a kulcskifejezéseket cseréltem le, így jött létre a Négy-Típusos Autonómiafelosztás (4TA) nevű séma, melyben a „szélsőhelyzeti autonómia" is helyet kapott. A gépi autonómia szintjeinek hivatalos elnevezése alapján világossá vált, hogy ezeket a technológiákat a szakemberek is csupán „fejlett automatikának" tekintik, és inkább a bulvársajtó terjeszti az *autonómia* szót, pl. az önvezető autók szintjeire. Ez

[129] 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.

főleg P1 terminológiai probléma megválaszolásához bizonyult hasznos megállapításnak. Vizsgálatomban a hivatalos felosztásnál cizelláltabb, általánosabban használható és sok fokozatú felosztásra tettem javaslatot. Ezután az első négy alfejezet gondolatai alapján összevetettem a kétféle autonómiát, pontokba szedve a gépi és az emberi autonómia főbb különbségeit. Ez a P2 védelmi probléma tekintetében lesz hasznosítható.

P1-gyel és P2-vel egyaránt kapcsolatos következtetések

R3.1: Kardinális különbségek azonosíthatóak az emberi és gépi autonómia között (III.5.).

R3.2: Az emberi és gépi autonómia közötti elvi hasonlóság, hogy a világ leképezését nem lehet objektivizálni (III.5.).

R3.3: A R3.1 és R3.2 alapján egy fejlett erkölcsi érzékkel rendelkező ember szabadsága egy gépben nem utánozható le.

R3.4: A gépeknél a beépített MI szintjéhez kapcsolódó autonómiaszintek meghatározásával lehet a különféle rendszerek besorolását biztosítani (III.4.1.); emberek esetében viszont az autonómia szintjei etikai alapon fejthetőek vissza (III.1-2.).

R3.5: Az R3.1, R3.2 alapján és R3.4. (III.4.1.) szintjei szerint, a jelenlegi vékony MI is csupán egy fejlett automatika, – ezt az önvezető autók hivatalos besorolása (III.3.2.) is alátámasztja.

P1-gyel kapcsolatos következtetés:

R3.6. Az R3.5 következtében az „automatika” szó megjelenítése egy MI-definícióban világosabbá teszi a kifejezést.

P2-vel kapcsolatos következtetések:

R3.7. Az emberi autonómiát etikai szempontból a szimplifikáló szint és a szélsőhelyzeti döntések szabadsága között szükséges skálázni (III.2.).

R3.8. A fejlettebb autonómiaszintet megvalósítani képes gépeknél a fehérlistásan meghatározott zárt működési szféra lehet csak képes a biztonságot garantálni (III.4.2.).

R3.9. Magas rendelkezésre állású rendszerek (HAS) legfeljebb delta szintű, azaz magára hagyható komplex autonómiával valósíthatóak meg („gépi user error” nélkül) (III.4.3.).

IV. A NEURÁLIS MI BIZTONSÁGI KIHÍVÁSAI

Mind a védelmi, mind pedig a terminológiai probléma (P2 és P1) szempontjából elkerülhetetlen annak áttekintése, hogy a neurális MI hogyan tekintendő újszerűnek, azaz milyen kihívásokat hoz – ebből a C1 és C2 célhoz is számítok részeredményekre. A kihívásoknak csak két szeletét vizsgálom a K4 kérdéskör[130] alapján: az egyik a technológia, a másik az emberek MI-hez való kapcsolódásának szemszögéből végez rövid áttekintést. A technikai áttekintést a kihívások listájával kezdem, ebben az első fejezetek információira támaszkodik (IV.1.). Ezután – folytatva az előző fejezet autonómiáról felvetett gondolatait – a gépi szabadság rossz kivitelezését vizsgálom meg, azaz a neurális torzítás (elfogultság, előítéletesség) áttekintése következik (IV.2.). Erre alapozva a két féle technológia tévedéseit összevetve elemzem, hogy mennyiben újszerű a gépek bemutatott torzítása a hagyományos számítástechnika programhibáihoz képest (IV.3.). A humán aspektusok áttekintésén belül először bemutatom azokat a buktatókat, melyek már okoztak visszaesést az MI fejlődésében (IV.4.), végül ezek közül külön alfejezetben emelem ki azt a fontos aspektust, mely arra mutat rá, hogy az emberek miért követik nehezen az ekkora szemléletváltással járó újdonságokat (IV.5.).

IV.1. A NEURÁLIS MI ÚJSZERŰ VONÁSAI ÉS EZEK KIHÍVÁSAI

Bár az eddigi fejezetek több oldalról felvillantották az MI újszerűségét, érdemes tematikusan is foglalkozni ezzel, mindkét kutatási probléma érdekében. Kimerítő leírás erről az újszerűségről nem adható – meg sem kísérlem, – még túlságosan az elején vagyunk az MI-korszaknak. Ebben az alfejezetben a kutatáshoz fontos technikai újdonságok jellemzését tűztem ki, melyek majd irányt tudnak mutatni a további vizsgálódásaimban. Ezek egybeesnek azokkal a tényezőkkel, melyek a technológia főbb kihívásait jelentik. A leírt jellemzők határai nem élesek, össze is függenek egymással, de közös az alábbi újdonságokban (kihívásokban), hogy igazi megoldásuk nincs, legfeljebb kezelhetőek kisebb-nagyobb kompromisszumokkal. Némely esetben elvi akadályai is vannak a tökéletes megoldásnak.

IV.1.1. Az MI újdonságának lényege az emberi mivoltunkat érinti

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,

[130] K4: A kihívások és a fejlődés kérdésköre: *Milyen kihívásokat hoz a neurális MI, hogyan kezelhető a hibázása, és milyen tényezők befolyásolhatják a technológia fejlődését vagy terjedését?*

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 emelőrúd 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 a kognitív képességek gépiesítésének perspektíváit még felfogni is nehéz, legújabb teremtményeink másképp újak, mint a mérhető és objektív külvilágban segítő gépek és műszerek. A novum nagysága épp az, hogy az MI olyan területeken képes segítségünkre lenni, amelyekről évezredeken át tűnt egyértelműnek, hogy az ember lényegét adják, így csak ember lehet rá képes, – sem állatok, sem emberi találmányok nem. Ebből az újszerűségből táplálkozhat az emberiség spontán félelme az MI-től, melyet erősít, hogy divatos téma a művészetekben, filozófiában.[131]

Emberi mivoltunk gépi támogatása csak úgy lehetséges, illetve csak úgy hatékony, ha a gép minél jobban idomul az emberhez a megismerés téren. Ez az idomulás több, mint ahogyan pl. egy szoftver felhasználóbarát, vagy amint ergonomikus egér idomul kezünk adottságaihoz. Ehhez egyfajta „lelki formatervezés" kell, aminek hatékony működése megoldhatatlan anélkül, hogy jobban megismernénk saját intellektusunkat, érzelmeinket, lelkünket, tudatunkat, akaratunkat, szociális dimenzióinkat, sőt morális érzékünket (és hasonló emberi tényezőket). Mindez a tudományok számára is új megközelítést hoz. Megvizsgálandó pl., hogy a saját kutatásaikon felül a kognitív tudományok gyűjtőfogalmába hogyan kapcsolódnak bele (V.3.1.).

IV.1.2. Az MI objektív mérhetősége és a Turing-teszt hibája

Röviden: az összetettebb MI-rendszerek már működési elveikben messze túlmutatnak a reáltudományok objektív mérhetőségének keretein. Tehát ez egy elvileg megoldhatatlan probléma, ráadásul két oldalról az. Elsősorban azért, mert az ember kognitív belső tere, amit az MI utánoz, nem mérhető olyan pontossággal, mint pl. a súrlódás. Bizonyos jelenségek tüneti oldalait tudjuk mérni – de egy ilyen mért okozat mögött a pontos okokra csak következtethetünk. A vizsgált alany egyedibb, összetettebb, változékonyabb, mint pl. az anyag vagy a hullámjelenségek. Másodsorban azért, mivel az utánzás mesterséges neuronháló általi megvalósítása is egy feketedoboz (ld. alább). Tehát amikor azt szeretnénk mérni, hogy a gép mennyire emberi, akkor sem az emberi oldal nem elég objektív, *amihez* hasonlítunk, sem a gépi, *amit* hasonlítunk.

[131] Olyan elképzelésekre gondolok, melyben az MI az evolúció új szintjét jelenti, új létrendet látnak benne, mivel az emberi felsőbbrendűséget szerintük létével cáfolja, pl. [27].

De még ennél is szubjektívebb a helyzet az első oldalon, mivel nincs olyan, hogy „az ember kogníciója”: az emberek kognitív képességei, tárgyi és gyakorlati tudása meglehetősen eltérő. Vagyis a híres Turing-teszt[132] elvi hibája, hogy „mérőműszere nem hitelesített”: hiszen azt sem tudjuk, mennyire tehetséges és mennyire rutinos (képzett) az a személy, aki a teszt során meg akarja különböztetni a gépi válaszokat az emberitől. Mondhatjuk, hogy egy IQ-teszt alapján bizonyos tartományba eső emberre értjük, – de nincs egy, a mérleghez vagy a Geiger-Müller számlálóhoz[133] hasonló, a fizikai objektivitás erejével bíró mérőmódszerünk sem erre, sem sok más MI körüli tényező mérésére.

A kognitív tudományok eredményeinek figyelembevételével képes a gép egyre emberibbé válni, sőt a gép hatékonysági mutatói is nagymértékben tudnak általuk javulni (V.3.1.). Jelen helyzetben ugyanis olyan újfajta kérdésekkel szembesülünk, melyek az eddigi gépeknél elvileg sem merültek fel. Pl. hogyan értelmezzük elvont kifejezések információtartalmát, főleg olyanokét, melyek a világszemlélettől függően fontosak vagy jelentéktelenek. (Pl. hősiesség, árulás, becsület, gőg, alázat stb.) Objektivizálhatóak-e, hogyan mérhetőek az ilyen címkékhez tartozó képzetek? Márpedig az MI viharos terjedése mellett egyre több ilyen kérdés merülhet fel. Ezeket a nyelvi modellek persze kezelik, pl. egymás mellett sorolnak fel néhány értelmezést. Ám ez máshol – pl. döntéshozó rendszerek számára, – kevés. A gépek emberiesedésével számos emberi tulajdonság mérésének igénye merülhet fel a gépek szintekbe sorolása és emberekkel való összevethetősége érdekében (III.4-5.).

Ez a fő motivációja annak a konvergenciának, mely által az MI-technológia beépíti magába a humántudomány számos eredményét, így a mérnökökön túl a fejlesztések szerves részévé válnak a humán tudósok is: nyelvészek, pszichológusok, szociológusok stb (V.3.). Ez túlmutat azon, hogy a számítástechnikai célrendszerek esetében a fejlesztésben eddig is részt vettek a célzott gazdasági, egészségügyi vagy állami terület szakemberei, jogászok, gazdasági szakemberek stb. Az újdonság az, 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). Korunkig a külső tudományokat inkább csak inspirációnak használta a számítástechnika, a hangsúly a természetes dolgok működésének algoritmizálásán volt (II.3.3.). Tehát az MI-ben a reál- és a humántudományok eredményei adódnak össze.

[132] A teszt azzal méri a gépek képességeit, hogy a tesztelő meg tudja-e különböztetni, hogy kérdéseire gép vagy ember válaszolt-e? Az eredeti teszt: [43].

[133] A radioaktivitás érzékelésére és számlálására használatos berendezés. https://www.elektroncspp.hu/cikkek/gmcspp.php (Letöltve: 2024.01.20.)

Ám ezen túl fellép egy lehatároló hatás is: A kognitív tudományok segíthetnek elválasztani, megkülönböztetni a gép lehetőségeit az emberi lehetőségektől A reáltudományok nem kompetensek megválaszolni például olyan kérdéseket, hogy mi az *értéke* az MI-jellegű emberutánzásnak? Vagy, hogy ez hogyan különbözik az emberi képességeinkből fakadó eredményektől? 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.

IV.1.3. A feketedoboz

Az előző fejezetben utaltam a neurális MI objektív mérhetőségének korlátaira. Ez a terület az emberinél sokkal jobban objektivizálható: a kód szintjén, az elvárt érték, vagyis a költségfüggvény minimumának megkeresése objektív (II.1.2.). Viszont abban mindig van némi szubjektivitás, hogy mit várunk el és milyen adatokat tudunk a gépnek adni (erre visszatérek, IV.2.). Itt azokat a problémákat vizsgálom, melyek a többrétegű neuronhálók „feketeségének" elvi velejárói. Ezeket (még) nem lehet teljesen 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.

- A feketedoboz 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 elrejti a neuronháló. Az (1) eset a bizonytalansággal (*uncertainty*) függ össze, és a következő pontba került. Az adatok problémája (2) alatt azt értem, hogy az MI-szolgáltatások felhasználói nem válogathatják meg a tanítóhalmazt, sőt nem is tudhatják pontosan, hogy mivel tanították be, vagy mivel frissítik az adott gépet. Ez védelmi szempontból is fontos probléma. Az átláthatóságot egyes fejlesztők blokklánc alkalmazásával próbálják elérni[134] – ez a javaslat sok esetben jó lehet, életszerűsége, terjeszthetősége azonban kérdéses.
- A feketedoboz **bizonytalansága** (kiszámíthatatlanság). A neuronhálók neuronjainak tartalma önmagában egy értelmezhetetlen számérték. Ezek megtekintése arra hasonítana, amikor egy futtatható fájlt (egy, amúgy működő programot) szövegszerkesztőben nyitunk meg, és csak egy értelmezhetetlen karakterhalmazt látunk. A legtöbb esetben ez nem jelent gondot, mivel az eredmény számít – egymás agyába sem látunk bele, csak azt tudjuk, amit a

[134] Ilyen pl. az IBM és a Casper Labs megoldása: [150].

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 életeket veszélyeztet. Gépektől még kevésbé fogadunk el ilyen kiszámíthatatlanságot. A programozott gépek szabálykövetőek, ráadásul szabályaik, az algoritmusuk jól átlátható. Ezért akár már emberéleteket bízunk automatákra; döntésük biztossága (és biztonsága) alapkövetelménnyé vált. Ezt a neurális MI a megszokott módon nem képes produkálni, mivel a valóságot nem pontosan ismert módon képezi le és egyszerűsíti le, ezért 0 számos felhasználás esetében nem minősíthetőek biztonságosnak az MI statisztikai alapú, bizonytalan döntései. Mérhető és részben kezelhető ez a gond [151], és kiemelten jelentkezik az egészségügyben [152], vagy a fegyverek (elsősorban az autonóm fegyverrendszerek) területén, és minden védelmi szempontból is kritikus rendszernél (erőművek, veszélyes üzemek). A komplexebb rendszerekben még nagyobb problémát okoz (pl. okosváros vagy még egy okosirodaház esetében).

- **Megmagyarázhatóság és értelmezhetőség**: ezek is kulcskifejezések a feketedoboz problémában (a Big Datánál is, bár ott kicsit másképp). Erre ugyan vannak megoldási javaslatok, de komoly kompromisszumokat igényelnek. Néha segíthet a modell egyszerűsítése. Más esetekben a post hoc értelmezés kínálkozik megoldásnak, ami azt takarja, hogy a feketedoboz ugyan megmarad, viszont betekintést próbálunk szerezni a kész modell eredményeibe (tehát gyakorlatilag egy átfogó és alapos tesztelésről van szó). Egy külön MI fejlesztési trend az XAI (Explainable AI), vagyis a magyarázható MI [153], melyre egyre jelentősebb a kereslet, annak ellenére, hogy az ilyen rendszerek hatásfoka lemarad a feketedobozzal dolgozó megoldásokétól. A probléma kezelésére irányuló számos bíztató törekvést figyelembe véve úgy vélem, ezek jelentősen nem fogják vissza az MI fejlődését, de terjedését akadályozhatják a megoldásokkal együtt járó komoly kompromisszumok.
- **A feketedoboz egyéb hátrányosságai**. Csupán felsorolásszerűen néhány további aggály, melyeket kutatók említenek. (1) Hibakeresés: nem kívánt eredmény esetén nehéz a javítás, nem lehet a bug-ot úgy lokalizálni és javítani, 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 a fals információkon keresztüli támadásokra.[154] (4) Bizalom, (5) elszámoltathatóság, (6) szabályozhatóság.

A fenti lista nem teljes körű, ám ennyi elegendő annak bemutatásához, hogy az MI rejtélyes belseje egy fontos hátráltató tényező mind az MI *fejlődésének*, mind pedig *terjedését* illetően.

IV.1.4. 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 részlegesek. Az AGI[135] megvalósíthatóságával kapcsolatban (ld. IV.6.) hasznos adalékokat ad, ha röviden megnézzük az említett három kulcstényezővel kapcsolatos helyzetet.

- **Általánosítás**. Emberi gondolkodásunk sarokkövei az egyre tágabb általánosítások (vö. IV.2.2.). A korai megerősítéses modellek nagyon rosszul kezelték a megtanult séma kis megváltozását is,[136] de jelentősek az előrelépések e tekintetben. Háromféle akadályt tudtam szétválasztani e téren. (1) Ha képessé tesszük a gépünket valami nagyon általános mintaegyezés azonosítására, akkor komoly humánerő igénye van annak, hogy ellenőrizzük, hogy ez a valóságban is létezik-e vagy sem. (2) A gép által fellelt mintázatok akkor használhatóak, ha a gép egy sémán belül dolgozik, így az újdonság csak mintázatsémákon belül új. A sémát az embernek kell jól definiálnia. (3). Gépeink nem képesek olyan általánosításra, amit a *filozófiai elvonatkoztatás* kifejezéssel írnánk le, hiszen erre a matematika tudománya sem képes (bár vannak erre próbálkozások). Az általánosítás terén áttörésnek tekintik az ún zero-shot modelleket, hiszen ezek képesek nem-tanult dolgokat is felismerni,[137] ezért bizonyos, eddig ismeretlen mintázatokat is felismernek. Ez azonban nem igazi általánosítás, hiszen csak az adott sémán belül működik, ahol nekünk kell megmondanunk az összefüggést, vagyis, hogy létezik a csíkosság és a ló összekapcsolása (vö. IV.2.6.). Ha a gép kapcsol össze a világban össze nem tartozó dolgokat, az nem hasznos. Összefoglalva: a gép általánosító képessége elvileg is korlátos – ezért is képtelen a *megértésre*.
- **Érvelés**. Az előző pont gondolatát folytatva: az olyanfajta érvelés, amely megértésen alapul, elvileg elérhetetlen egy mostani mélytanulási modell számára. Erre az iménti általánosítással összefüggő korlátosságon túl több érvet is felhoznak a kutatók[138], melyet aránytalan lenne itt elemezni. Szemléltetésül egyet vázolok: a virtuális térbe mindig csupán egy beál-

[135] AGI = Artifitial General Intelligence (ld. I.3.3.).

[136] Például egy régi számítógépes játékot ügyesen játszó egyszerű mélytanulás az ütő minimális megváltoztatása után nagyon rosszul teljesített.[155]

[137] 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.

[138] Pl. hosszú távú tervezés, algoritmikus adatmanipuláció. Bővebben ld. [156, pp. Section 2 of Chapter 9].

lított világnak a részhalmaza kerül leképezésre, ahol az összefüggések is csak tanult információk, próbálkozások, nem pedig valódi felismerések. Vagyis a gép csak adott érvelési sémát képes követni – ami elegendő arra, hogy pl. egy lehetséges stabil molekula létét megjósolja, így rengeteg kísérletet spórolva meg. Igazi érvelésre ugyan nem képes, viszont valójában nincs is szükség arra, hogy a gép tényleg megértsen egy érvelést, elegendő, ha jól utánozza azt. Tehát a problémát manapság inkább megkerülik, mintsem megoldják. Ez a cél jobban megvalósítható, és számos esetben éppúgy elegendő, ahogyan egy magolós gyerek is kaphat jelest a feleletére, annak ellenére, hogy nem érti a tananyagot. Ha nem faggatja nagyon a tanár, képes ügyesen úgy tenni, mintha értene is valamit abból, amit felmond. Jó példák erre az NLP-rendszerek, 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 feladattípusnál új módszer kell, és még így is csupán célrendszerek jönnek létre. Így a mai eredmények aligha vethetőek össze egy felnőtt ember önképzésével, melyben a mintakezelésen 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 és meglehetősen diszruptív tudáspotenciál.

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 a jelenleg ismert fejlesztési irányokban megugorható lenne.

IV.1.5. A nyers erő paradigma kérdőjelei

A jelenlegi MI-fejlesztések a problémákat *brute force*[139] megoldásokkal közelítik meg, miközben a „lemásolt” agy mindössze húsz watt energia felhasználásával sokkal több mindenre képes, mint egy óriási erőforrás-igényű, fejlett MI-rendszer.[157] Ez a tény önmagában is rávilágít arra, hogy mennyire megkérdőjelezhető a jelenlegi fő paradigma, mely a hatalmas adatmennyiséggel és a szupergépek teljesítményével képes csak jó eredményekre.

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

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 is.[157] 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 az emberéhez hasonló „húsz wattos megoldást".

IV.2. AZ MI TORZÍTÁSA (ELFOGULTSÁG) – AVAGY EGY RÉGI PROBLÉMA ÚJDONSÁGAI

Az előző alfejezet listája is érintette ezt a kérdést, melyre itt érdemes alaposabban kitérni. A torzítás lehet nem szándékos vagy egy MI-szabotázs. Itt az MI-torzítás *tudatos elkerülésére* koncentrálok, csak néhol érintem a *tudatos torzítás* témáját (ami a torzulás hátterének leírása közben úgyis kézenfekvővé válik).

IV.2.1. Az MI-torzítás terminológiai vizsgálata és felosztása

Népszerű a témát olyan kifejezésekkel tárgyalni, mint „az MI előítéletessége", az „MI elfogultsága", én azonban a torzítás kifejezést preferálom. Ennek oka kettős: egyrészt mert általánosabb, jobban lefedi a hibatípusok sokféleségét, másrészt objektívebb, kevésbé sugalmazza azt, hogy egy gépben hasonló ez a jelenség, mint az emberben.

Az előítéletesség szó ilyen használata ellen két érv hozható fel. Egyrészt az ilyen vádak sokszor nem a rendszer hibáját tükrözik, hanem a vádoló nem akar szembenézni az objektív tényekkel. Ha az uránbányászok betegségei között alulreprezentált a méhnyakrák a társadalmi átlaghoz képest, az vélhetően nem hiba, hanem a bányászok születési neméhez köthető tény. Másrészt az *előítéletesség* szónak van pejoratív felhangja, mely egyfajta tudatosságot sugall, mint ahogyan egy neonáci dönt a politikai ideológiája mellett. Ennek a szónak a használata pont azoknál jellemző, ahol amúgy is rettegnek a tudatra ébredő, ember ellen forduló MI-től. Vagyis az *előítéletesség* kifejezés helytelenül antropomorfizálja a jelenlegi, vékony MI-modellekkel

működő gépeket. Emellett a fejlesztők számára is sértő annak sugallása, hogy úgy teszik előítéletessé az MI-t, ahogyan egy szülő vagy tanár befolyásol egy fiatalt. Tehát ilyen esetekre csak az *elfogultság* kifejezést tartom helyesnek, mely emberekre alkalmazva is jobban magába foglalja a nem szándékolt hátteret. Hasznos lesz számunka még a következő elemzésnél az előítéletesség szó, de csak ezt a szétbontott alakot etimológizálva.

Tehát a szóhasználat terén egy világosabb terminológia mellett szállok síkra, ahol a fejlesztői hibákat MI-torzításnak, az MI-k működésének szándékos befolyásolását pedig MI-szabotázsnak hívják. Hiszen valójában a kérdéskör tudományos vizsgálata sem a rossz szándékból elkövetett torzításra fókuszál, hanem a rendszer bonyolultságából adódó, ilyen jelenséget okozó problémakört elemzi. Tény ugyanis, hogy a jó szándék ellenére is számos esetben le kellett állítani MI-ket a torzítás miatt, – viszont ennek csak azok az esetei lettek a világsajtó témái, melyek az emberi előítéletesség sémái szerint voltak hibásak. Elsősorban a faji és nemi téren figyeltek fel erre [158], például az MI-re alapozott munkaerő-toborzás nemek szerint torzított, és hírhedtté vált az az eset, amikor egy chatbot lett rasszista.[159] S bár hasonló hiba okozza, de nincs sajtóvisszhangja pl. egy olyan üzleti szempontból problémás esetnek, amikor egy mélytanulást is használó webshopban az automatikus tanulás miatt egy egyedibb kinézetű cipő nem jelenik meg egy keresőkérdésre, holott megfelel a keresés feltételeinek (kézi címkézésnél megjelenne).

A torzítás okait célszerű a hibák kezelhetősége szerint csoportosítani. Sokféle lista fellelhető, én hét főbb tényezőre bontva fogom a kérdést elemezni (több közülük feloszthető):

0. **fejlesztői belső**: a létrehozó személyek emberi elfogultsága (ez a többire is hatással van);
1. **hardver**: alapvetően hagyományos hardveres hibák, de egy hardveres neurális megoldás alkalmazása is lehet hibás;
2. **viszony**: máshol bevált technika (modell, adatválogatás vagy betanítás) helytelen implementálása;
3. **algoritmus**: kódhiba, vagy a modell elégtelensége;
4. **adat**: az adathalmaz kiegyenlítetlen, vagy a címkehalmaz nem megfelelő;
5. **tanítás**: a tanítási metódus (vagy személy) elégtelen a súlyok elfogadható hangolására;
6. **felhasználói hiba**: alapvetően azt jelenti, hogy a megkapott eredmény értelmezője félreérti az MI-től kapott eredményt, de jelentheti a felhasználóktól tanuló MI torzulását is.

Külön kihívás, hogy az MI feketedoboz-jellege miatt utólagos tesztek során is nehéz felismerni a döntések torzítását, így ezt a hibatípust (is) csak preventíven lehet jól kezelni. Bár egyes

tényezők (3–5) torzításának elkerülése részben automatizálható,[140] de a legtöbb szakember egyetért abban, hogy a torzítások elleni prevenció alapja az MI egészének gondos létrehozása. Ezért vettem külön a 2. tényezőt, vagyis a 3–5 tényezők viszonyát. Többen rámutatnak arra is, hogy csak megfelelő emberi akarat adhat esélyt a probléma csökkentésére,[141] ezért a felosztásomban szereplő „nulladik tényezőt", vagyis az emberi szempontot most kiemelten tárgyalom.

IV.2.2. Fejlesztői hibák és a humánvirtualizáció alapproblémái

A gépben megjelenő torzítások bemutatása előtt szükséges a fenti lista nulladik pontjának vizsgálata, a létrehozók tudattalan elfogultságának, vagy arra vezető szemléletüknek vázlatos elemzése. (A témával foglalkoztam más összefüggésben is [162, pp. 369].)

(1.) A humánvirtualizáció és annak alapproblémái

Az emberi elfogultság fő oka a világ szükségszerű egyszerűsítése, erre a humánvirtualizáció fogalmat alkalmazom. Itt nem célom a kogníció folyamatának tudományos bemutatása, mivel ez túl messzire vezetne – csupán szeretném ebből a sajátos irányból megvilágítani belső (virtuális) világunk kialakulási mechanizmusának egyik elemét. Kognitív képességeink összessége hozza létre azt a belső értést (vagy félreértést), tudást, emléket vagy asszociációt, amelyre következtetéseinket, így döntéseinket visszavezethetjük. A humánvirtualizáció tehát a világ leképezése a személyes kogníciónk által alkotott világunkba, ennyiben párhuzamba állítható a gépi virtualizációval. A kibertér azonban nem egy standalone PC belső működése, így a személyes kognitív világunk a szociális hálóval és társadalmi hatásokkal együtt hozza létre azt, amit valójában kognitív térnek nevezhetünk. De maradjunk most a személyes síkon.

Mondhatjuk, hogy az emberi elfogultság azért keletkezik, mert a világot csak egyszerűsítve tudjuk megfelelően kezelni – ez a mondat is egy leegyszerűsítés. A végtelenül változatos világot kénytelenek vagyunk megszámlálható, kezelhető módozatra leképezni, ezért hozunk létre halmazokat, címkéket. A címkézés kognitív rendszerünk egyik alapja – részben ezért is alapoztuk erre a számítógépes rendszereinket, ezen belül az MI-ket is. Előnye, hogy eleve kezeli a kisebb eltéréseket az egyedek között (pl. alma), hierarchikusan is rendszerezhető (pl. az alma fajták szerint), de egy elem több címkével is ellátható, mely alapján tőle teljesen idegen dolgokkal is egy halmazba kerülhet a címkézett dolog (piros, gömbölyű, édes). Továbbá ez képes kezelni, kvantálttá és mérhetővé tenni az analóg jellegű világ tényezőit, hogy aztán kezelhető,

[140] Pl. az IBM ilyen rendszeréről [160]. A rendszer leírása: https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/

[141] Javasolt összefoglaló: [161].

számolható, objektív valóságot hozzon belőle létre. Más esetekben a címkék fokozatokat jelentenek (gyors-lassú), és épp a köztük való átmenetet kell trükkösen leképezni a gépbe (pl. fuzzy logikával, II.3.2.)

Az emberek csak a korábbi tapasztalataikból kialakított rendszerbe képesek az új információkat beletenni. A gép ezt utánozza, részben a hagyományos memóriáiban tárolt adatokkal, részben a mélytanuló modellek rövid és hosszabb távú memóriájával (II.2.2.), vagyis az MI is korábbi (részben emberi) tapasztalatokra alapozza tudását. Mindkét tudást a belső címke és halmazrendszer mintázatai generálják. A kogníció működésének ezt a részét jól ábrázolja az „előítéletesség" kifejezés etimológiája: ítéleteinket mindig korábbi ítéletek előzik meg. Tudásnak akkor érezzük, amikor ellenőriztük ezeket a korábbi ítéletünket, és az igazolódni látszott. (Bár az igazoltnak tűnő tudás is megdőlhet, amint azt a kvantumfizika paradigmaváltása mutatta.)

Kimondható tehát, hogy *kogníciónk lényegét tekintve előítéletes*. Vagyis az „előítélet-mentesség" kifejezés ismeretelméletileg értelmezhetetlen – morális szinten ettől még rámutathat arra, hogy bizonyos prekoncepciótípusokat célszerű, illetve etikus felülvizsgálni. Tudáshibát generál az egyszerűsítési ösztönünk túlzása is. Ennek esetei például (a teljesség igénye nélkül):

- túl kevés adat alapján jön létre a címkézés (felületes ismeret),
- túl kevés a címke (terminológiai keretek hiánya),
- egy címke alá való besorolás feltételrendszere szimplifikáltabb a kelleténél (a terminológia pontatlan ismerete),
- bizonyos címke indokolatlan súlyt kap (erre példa lehet a minden kutyától való félelem, amikor egyetlen agresszív kutya emléke torzítja el a sok szelíd kutya emlékeinek súlyát, így a kutyához az agresszív címke mindenképpen odatársul).

Kisgyermekeknél az ilyen tévedéseket természetesnek vesszük, hiszen bizonyos, hogy nem elégséges az adat és a címke, ami alapján általánosítani próbálnak. A fejlődés során azonban megjelenhet az új minták tudatos vagy tudat alatti elnyomása, figyelembe nem vétele, ezért felnőttek esetében az előítéletesség szó pejoratív alkalmazása morálisan indokolt lehet. A spontán (figyelmetlenségből, érzelmekből) generálódó emberi előítélet tehát nehezen kerülhető el. A tudatos előítéletesség (pl. valaki nem hajlandó vásárolni őstermelőktől a piacon) viszont végtelen viták tárgya lehet. A humánvirtualizáció ilyen sajátosságait nem kívánjuk implementálni intelligenciánk utánzásakor – azonban ez nemigen kerülhető el.

(2.) Emberi és gépi virtualizáció

Az analógia ellenére ki kell emelni, hogy az analóg világot bensőnk másképp egyszerűsíti le, mint a gépek, itt háromféle különbséget kívánok leírni. Először is abból indulok ki, hogy

mindenki a világ sajátos, szubjektív leképezését hozza létre magában. Ám a többi emberrel való interakció érdekében mindenki már gyerekként megtanulja objektivizálni ezt a belső világát, mert csak ilyen formában tudja közölhetővé tenni saját tapasztalatait és gondolatait, és befogadni másokét. Az objektivizáláshoz egy adott nyelv formáit, szokásokat (illemet), mértékegységeket és hasonlókat használunk, ezek által is tárolunk információnkat. De a világból nyert információk jó részét nem ilyen objektív módon tároljuk – ezért jelennek meg pl. a félreértések a kommunikációnkban. Az első eltérés a gépi és emberi tudás között abból adódik, hogy az objektivitás tényezőit is mi hoztuk létre, ezzel szemben a gépek az általunk létrehozott, már objektivizált információt vagy adatot tárolják.

A számítógépes kutatásokon belül külön szakterületek foglalkoznak az ilyen problémákkal: ún. tudásreprezentációkat, illetve ontológiákat dolgoznak ki. Ezek a régebbi, még a hagyományos számítástechnika alapjain kialakult leképezési rendszerek (V.1.1.), de az MI erősödése miatt reneszánszukat élik. Tudásreprezentációnak az olyan módszerek összességét hívják, amelyek lehetővé teszik számítógépes rendszerek számára a világ strukturált leírását. Ebbe számos módszer tartozik: a szimbólumok és logikai szabályok reprezentálásától (szimbolikus MI) a fogalmaknak és azok kapcsolatainak gráfos ábrázolásán át (szemantikus hálók) egy feladat összes releváns információját leíró struktúrákig (keretrendszerek).[163] Az ontológiák pedig egy adott tartományban előforduló fogalmakat definiálnak, és leírják azok tulajdonságait és a közöttük fennálló kapcsolatokat. Az általuk biztosított közös szókészlet és nyelvtan teszi lehetővé az adatok szemantikai leírását és a tudás megőrzését a gépi reprezentációban.[164] Ezek a tudományok tehát olyan leképezéseket hoznak létre, melyek matematikailag ragadják meg a világ adott szegmensét – ám egy ilyen leképezés is csak önmagán belül képes objektív lenni. Például az „elhanyagolt tényezők” határát egy adott feladatnál az igények és tapasztalatok alapján tudják csak meghatározni.

A második lényeges eltérés, hogy mi emberek, tudat alatt is folyamatosan átdolgozzuk a tárolt dolgokat: próbáljuk kipótolni a hiányzó információt, összefüggéseket találni, és érzelmeket társítunk hozzájuk. Ezt az utótorzítást néha hibaként, néha tudományos felismerésként vagy művészi látásmódként lehet értékelni – de mindenképp „emberi”, tökéletlenségünk összes szépségével együtt. Példák erre azok a tudományos elméletek, melyek sokszor hosszú ideig meghatározóak, objektívek voltak, míg egy másik modell nem adott pontosabb közelítést a problémára. A fejlettebb gépek is képesek arra, hogy folyamatosan átdolgozzák az adataikat, de ezt nem ösztönösen teszik, hanem csak létrehozóik igényeinek mértékében.

A harmadik különbség a tudattalan szemlélet hatása a döntésekre. Az előző fejtegetések alapján világossá válhatott, hogy a döntéseink mögött a tudattalan szemléletek jelentik a legnagyobb kihívást, mivel ezek tudatosítása igen nehéz és az egyén számára személyes, önismereti feladat. Könnyen hozhat valaki az explicit elveivel ellenkező elhatározást, hiszen a vallott etikai elveink mellett pl. tudatalatti tényezők is befolyásolják az emberi döntéseket.[142]. Ezzel szemben egy gép esetében a probléma épp ennek fordítottja: a világos parancsokat a gép túl könnyen végrehajtja; egy gépben nem lépnek fel tudatalatti vagy lelkiismereti gátak. Az autonóm fegyverek elleni averzió egyik fő oka, hogy a gépnek teljesen egyforma output feladat az emberölés vagy egy kavics arrébb rakása. Kérdéses persze a parancs világossága, illetve a bonyolódó rendszerekben a neurális hálóban létrejövő „gépi tudatalatti" szféra (vagyis az MI-torzítások) hatása. Mind az MI kialakításánál, mind a kiképzés során a tudattalan részek elkerülésére törekszenek, ám a gépeknél ez talán nem annyira nehéz, mint az embereknél.

(3.) A tudattalan szemléletek hatása és ennek védelmi aspektusa

Az előző lábjegyzetben hozott példa jól mutat rá arra, hogy egy MI-rendszer megalkotója hogyan csempészheti bele elfojtott vagy teljesen tudattalan szemléletét élete minden területébe, esetünkben az MI létrehozásának folyamatába. Hiszen az ember elvei nem csupán számtalan ponton megfogalmazatlanok, hanem sokszor következetlenek, valamint változékonyak is. Érzékeinket, érzékelésünket is meg lehet zavarni, ezek is hatnak, továbbá a személyes környezeti ingerek (elvárások, érzelmi állapot) vagy akár a média felől jövő óriási mennyiségű hatás összességét sem tudjuk feldolgozni, jól rendszerezni, megfelelően elhelyezni a korábbiakkal összefüggésben. Régen is keveseknek sikerült, de korukban egyre nehezebb következetes és tudatos nézetek szerint élni úgy, hogy ráadásul ezek az alapszemléletek ne avuljanak el néhány év alatt. Egy konzekvens szemléleti tudatosság megvalósítása igen nagy személyes kihívás, melyre kevesen vállalkoznak. Erre vállalkozni csak belső, lelki motiváció alapján lehet – pedig erre a gépi intellektus fejlesztésekor is egyre nagyobb szükség lenne.

A szemléleti torzítás már a modellnél vagy az algoritmus szintjén is belekerülhet egy kódba, hiszen a jobb programozóknak stílusuk van – de az adatok meghatározása vagy az emberileg felügyelt tanítás során elkerülhetetlen az ilyen jellegű „figyelmetlenség". A szó azért került idézőjelbe, mivel nem szétszórtságot értek alatta, hanem azt, hogy az ember saját szemléletének

[142] Például igen sok katona, – bár hivatása, hogy hazája védelmében akár embereket is öljön, – szereti a sülthúst, mégsem bírja elviselni egy állat levágását, sokan az arról való beszédet sem. A két dolog közötti ellentmondást egyszerűen nem tudatosítják, sőt elfojtják, kizárják a tudatosságukból ezt a kellemetlen összefüggést. Ennek egy harci helyzetben bajtársaik megvédésének elmulasztása lehet az ára. (Megjegyzés: erre a katonák pszichéjét tréningezni kell, pl. haszonállatok megölésével, amely egyben túlélési gyakorlat is).

foglya, így önnön nem tudatos részét képtelen lesz kizárni a munkájából, észre sem veszi, amikor belecsempészi.

A mai számítógépes rendszereket nem egyének hozzák létre, ám az imént leírtak igazak a csapatmunkára is. Egy fejlesztő csapatnak is van egy sajátos arculata. Ideális esetben „egyre jár az agyuk", ami ad egyfajta virtuális személyiséget a közösségnek. Ez hatékonnyá teszi a közös munkát, ám épp ezért átsiklanak esetleg egy-egy probléma fölött, melyet az egyre járó agyak egyike sem vesz észre. Közös szemléletüket (a virtuális személyiséget) meghatározhatják a vezéregyéniségek vagy esetleg a vezetők, de óhatatlanul kihat rá a kultúrkör vagy szubkultúra is, ahol szerveződik.

Ezért lehet érzékelhető eltérés MI-fejlesztések között akár egy államon vagy kultúrkörön belül is, de még nagyobb különbségre lehet számítani pl. egy kínai, egy orosz, egy amerikai vagy egy iráni team esetében. Ezek a teamek köznapi kérdésekben együtt gondolkodnak, trivialitásaik vannak – melyek azonban egy másik kultúrkör számára furcsa szabályként jelenhetnek meg. Az amerikai nyelvi modellek egyeduralma miatt egyelőre ennek hatásai nem érzékelhetőek és még nem jól kutathatóak, azonban a kapuban vannak már más modellek is, ahol ilyen jelenségek véleményem szerint meg fognak jelenni.

Amennyiben ezt a problémát ki akarják küszöbölni, akkor jó megközelítés egy tudatosan nemzetközileg, online szerveződő team: ebben a tagok egyedi szemléletének különbségei segíthetnek tudatosítani vagy észrevenni az előbb említett eldugott szemantikai vagy metodikai hibákat, így egy kevésbé elfogult rendszert kapnak. De a kognitív eltérések tudatos keresése ki is használható úgy, hogy az imént említett eltéréseket nem feloldják, hanem az ellenfél rendszerében támadják. Ez a gondolat átvezetne a kognitív hadszíntér témájához – mely egy még most meginduló terület, kevésbé kiforrott, mint az MI, – de ne térjünk el a tárgytól.[143]

IV.2.3. Kitérő: az Üres Lapra alapuló objektív MI kritikája

Az imént humánvirtualizációnak nevezett probléma elkerülésére felmerül az emberi torzítások elkerülésének egy olyan módja, hogy a gépeket nem kényszerítjük rá azon címkék és halmazok alkalmazására, amelyeket mi emberek tanultunk meg és adunk nekik. Ennek a megközelítésnek egyik szélsőséges felvetése szerint rá kellene hagyni egy ilyen MI-re, hogy maga alakítsa ki az input adatok értelmezésének miként jét. Egy ilyen abszolút objektivitást megcélzó

[143] Az itt vázoltakhoz még annyit érdemes megjegyzem, hogy a metaverzumok vizsgálatával is igazán érdekes további összefüggések tárhatóak fel az imént leírtakkal összefüggésben.

kutatói koncepció képviselői szerint minden emberi adat szubjektív, ezáltal valójában akadályozzák egy erős mesterséges intelligencia optimális képességeit. Az egyik szerző, – általánosítva azt, hogy az arcfelismerési algoritmusok nem használhatóak más célra, – arra a következtetésre jut, hogy „a világról az MI-rendszer által megfigyelt priori információk felhasználása nemcsak a kifejlesztett numerikus módszerek alkalmazhatóságát korlátozza, hanem teljesen lehetetlenné teszi egy erős mesterségesintelligencia-rendszer létrehozását is."[144] [165, pp. 6] Tanulmányában kimutatja, hogy egy megfelelő szenzorrendszer-hálózattal ellátott, de „üres lappal" induló MI lehet majd képes újszerű problémák megoldására. A tanulmányból nem egyértelmű, hogy mennyire „általános" (és nem vékony) MI jönne létre, sem az, hogy 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 ezen hiányosságokat félretéve, a teljesen önmagát kialakító MI tézisével én elvileg sem tudok egyet érteni, a következők miatt:

- ez a „üres lap teória" nagy valószínűséggel egy ember által érthetetlen, sajátos kommunikációt hozna létre, tehát értelmezhetetlen halmazokat, címkéket és összefüggéseket. Vagyis az embernek kellene megfejtenie és megtanulnia az ilyen MI által létrehozott saját nyelvet, aminek a logikája eltérne minden emberi nyelvétől.
- Még csak esélye sem lenne az embernek megtanulni „üreslapMI-ül". Ugyanis ez a saját leképezés a világ változásával valós időben változna, tehát sokkal gyorsabban, mint azt az emberek követni tudnák. Egy ilyen modell nyelvi oldalán a kifejezések folyamatosan átalakulnának, a régiek érvénytelenné válnának, alkategóriákra osztódnának. De az ilyen rendszer egyéb (nem nyelvi) tudását is egy időpontban rögzíttetni kellene vele, hogy esély legyen az értelmezésére, hogy abból valami számunkra hasznos eredményre jussunk.
- Bár a cél a humán előítéletek elkerülése volt, a jelenlegieknél sokkal veszélyesebb torzítást hozna létre az ilyen MI. Ugyanis az ember számára érthetetlen, új torzításmódokat eleve esélytelen kezelni, mivel ismeretlenek, de elkerülésüket nem tudjuk garantálni.

Ezek alapján egy ilyen „üres lap" által önmagát kialakító MI igen veszélyes, mivel a döntései mögötti leképezések nem emberiek, így döntéseitől sem várható, hogy emberi igényeknek és elvárások szerint jönnek létre. Lehetnek optimálisak (a gép szerint), de szinte bizonyos, hogy mindenféle etikai alapelvet teljesen nélkülöző döntések lennének. Mivel a gép nem él, ezért az

[144] Saját fordítás. Megjegyzendő, hogy a cikk írója orosz, így ezzel a radikális gondolattal rámutat az eltérő kulturális háttérből adódó eltérő tervezői szemléletre.

életet például miért tisztelné? Ezért csakis biztonságos, szimulált környezetben vethető fel egy ilyen kutatási célú vizsgálat, és akkor is időben korlátozva kell tanítani az üres lapról, eredményeit pedig ember által érthető másik MI segítségével elemezni.

IV.2.4. Tényezők viszonya, algoritmikus és hardveres torzítás

Rátérve az MI-ben megjelenő elfogultsági problémákra először is hangsúlyozni kell, hogy a tanítható rendszerek torzításában a modell, az adatok és tanítási metódusok összefüggésben vannak, egymásból következnek. Így a tanító adatok és az emberi-automatikus tanítás aránya függ a modelltől is, nemcsak a feladattól, tehát a hiba nem csupán a fenti kategóriákból ered, hanem ezek egymáshoz rendeléséből is adódhat. Bár megkérdőjelezhető a részfeladatok viszonyának külön szintként való interpretálása, hiszen a fejlődés jelen állapotában nem ez szokott a probléma oka lenni. Azonban úgy vélem, hogy az egyre könnyebben használható tanító keretrendszerek terjedésével, újabb és újabb saját célra MI-t barkácsoló emberek vagy team-ek könnyen elkövethetik ezt a hibát.

Továbblépve az algoritmusokra: kimondható, hogy egy MI akkor torzít, ha rosszul egyszerűsíti a világot a vizsgált kérdéskörre. Ezt kiemelten nehéz észrevenni algoritmikus szinten, hiszen a matematika soha nem téved. Annak adott alkalmazása azonban lehet hibás, vagy nem optimális. Tehát az algoritmus szorosan összefügg a nulladik szint humán tényezőjével, a világ belső leképezésével (és annak hibáival) is. Algoritmikus hibák alatt meg kell különböztetni a kódolási és tervezési hibát. Az előbbit elegendő említeni, mivel megelőzése és kezelése hasonló a hagyományos kódhibákéval (erre visszatérek IV.3.-ban), ám a tervezési hibákról néhány gondolat még ide kívánkozik.

Amint az a modellek bemutatásánál (II.) világossá vált: egy-egy matematikai modell nem optimális minden kogníciótípusra (bemutattam pl. a képi és a szöveges kihívás közötti jelentős eltéréseket). Tehát a nem megfelelő modell választása nem feltétlenül „hibát” generál, viszont könnyen hozzájárulhat a rendszer torzításához. Egy megfelelő modell ügyetlen implementálása is hasonló következményekkel jár. Hiszen egy szemantikus hibát egy hagyományos kódban is nehéz észrevenni, főleg, ha tévedés miatt a tesztelésből is kimarad valamilyen, a tervező számára nem ismert kivételtípus vizsgálata.

Végül megemlítendő, hogy az MI-t futtató gépi architektúra szintjén is előfordulhatnak torzítást okozó hibák, a sérült memóriaszektortól kezdve a frissülési (adatkapcsolati) problémákig (akár a hagyományos programoknál). A jövő torzítási hibái azonban várhatóak a felhasznált architektúramodell téves ismeretéből is. Bemutattam (II.), hogy jelen hardverfejlesztési tendencia igyekszik egyre inkább az MI-rendszerek „alá dolgozni”, illetve akár fizikailag valósítanak

meg neurális modelleket. Ilyenkor a gépben is megjelenik egyfajta torzításra való hajlam. Pontosabban az ilyen hardverek helytelen használata[145] adhat nehezen észrevehető hibákat. Ez is hipotetikus problémának tűnik, de érdemes elképzelni olyan eseteket, amikor egy célhardverbe való befektetés nem térül meg az eredeti módon, akkor kézenfekvővé válik annak más, jövedelmezőbb célra való használatának erőltetése.

IV.2.5. Az adathiba

Egy érdekes kutatás, „a pszichopata MI" bemutatásával vezetem fel a problémát. Egy kutatócsapat szándékosan „negatív" képekkel tanította az egyik kísérleti MI-t, míg egy azzal azonos modell pozitív képeken tanult. Ugyanaz az algoritmus így nyilván eltérő eredményre jutott. Ennek ellenőrzésére egy pszichológiai tesztet oldattak meg velük: az egyik egy pszichopata reakcióit adta, míg a másik MI egy normális emberhez hasonlóan reagált. Ebből a kutatás vezetője arra jutott, hogy az adatok fontosabbak az algoritmusnál.[166][146] Megjegyzendő, hogy ez a következtetése annyiban sántít, hogy nem végzett összehasonlítást eltérő modellekkel: kérdés, hogy nem kapnánk-e hasonló eredményt ugyanazokkal az adatokkal tanítva egy rosszul beállított és egy jól működő modellt. Helyesebb lenne tehát úgy fogalmazni, hogy a jelenleg használt rendszereknél nem az algoritmikus szint a legveszélyesebb ok, ami egy-egy jelentősebb hatású torzítás mögött áll, sokkal nagyobb probléma a tanító, validáló és tesztelő adathalmazok megfelelő összeállítása. Az adathiba elkerülésénél félrevezetéstől mentesített adatokra törekednek, ezen a téren két kihívást választok szét.

Az adatokkal kapcsolatos kihívás egyik fele hasonlítható a kérdőívkészítés objektivitásra törekvésére, csak nem a megkérdezettek valósághű eloszlására tervezik meg az elvárásokat, hanem az adatokéra. Ez nem feltétlenül egyenlőséget jelent (bár van, amikor így értelmezik), hanem azt, hogy a valódi számarányuknak megfelelően jelenjenek meg a gép válaszai között olyan halmazok is, melyek spontán módon kevésbé reprezentáltak a digitális térben. Pl. az idős emberek kevesebb médiatartalmat gyártanak, ezért csak személyes (nem automatizálható) módszerekkel lehet tőlük adatot bekérni.

A kihívás másik fele a döntési adatok frissességének és hitelességének ellentmondásával kapcsolatos. Ha pl. egy nyelvi modellben az internet friss adatait is figyelni szeretnénk, akkor a következő lehetőségek adódnak. Az első, hogy gyakorlatilag pozitív cenzúra alá kell vetni

[145] Nyilván nem egy szövegre optimalizált rendszer képfeldolgozására való alkalmazását értem ez alatt, hanem pl. egy neurális processzor dokumentációjának pontatlan ismeretét.

[146] A kutatásnak weboldala is van: http://norman-ai.mit.edu/

annak felhasznált tartalmait: kiszűrni a hamis, a más MI-k által generált, vagy csak többszörös tartalmakat közlő forrásokat (feketelista). A második eshetőség, hogy a válaszhoz csak emberek által minősített (pl. tudományosság szempontjából releváns) tartalmakat venni figyelembe, ha ez elvárásként merül fel (fehér lista). Egy harmadik módszer, amit az általam is segédeszközként használt Perplexity rendszer is alkalmaz: a kettő keveréke (szürke lista). Pl. Academic mode-ban előnyben részesíti a tudományos portálról származó tartalmakat és kerüli (de nem zárja ki teljesen) az ennek a kritériumnak nem megfelelő, de relevánsnak tűnő tartalmakat. Persze számos egyéb szűrőt is használ, mint a közzététel ideje vagy a szerző hitelességének vizsgálata. Ez utóbbi tényező megállapítása külön érdekes, mivel tudatosan és indokoltan kerüli az olyan metrikákat, mint a Hirsch-index, inkább átláthatóság, semleges fogalmazásmód, publikációs előzmények stb. alapján maga ítéli meg a szerző hitelességét az automatika.[167][147] Ez a példa jól mutat rá a különböző igények vagy célrendszerek frissességének kihívásaira és azok kezelésére, de az is látszik belőle, hogy hosszú távon hogyan változtathatja meg az MI a tudományos munka mai metrikáit is (jó esetben pozitív irányban).

A nehézség vázolása alapján világossá vált, hogy az adatok megfelelőségének kérdése is kapcsolatos a nulladik szinttel, hiszen a tökéletlen emberi megismerésre alapozzuk. Például a kutatható adatok torzítása, egyoldalúsága miatt egy kutató is áldozatává, sőt részesévé válhat egy információs megtévesztő műveletnek, amikor internetes forrásokból tájékozódik. Igaz ez akkor is, amikor közvetlenül gépi (szenzor-) adatokat dolgoz fel az MI, hiszen azt, hogy ezeket hogyan vegye figyelembe, az emberi megismerés alapján alakítják ki a fejlesztők.

IV.2.6. Tanítási hiba és a felhasználói probléma

Alább nem részletezhetem azt, hogy hogyan lehet jó modellt jó adatokkal rosszul tanítani, csupán a hibatípus jellegét járom körbe. Az igazi kihívás az ember által felügyelt, a félig felügyelt és a teljesen automatikusan tanuló fázisok arányainak helyes meghatározása, hiszen ehhez illesztve kell az előbb elemzett adathalmazok összeállítását is megtervezni. Ide tartozik azoknak a módszereknek a kiválasztása is, melyekkel megfelelő tanítóadatokat generálunk: pl. hogyan szűrik őket, hogyan generálnak adatot (pl. a tanító kép milyen típusú változatait hozzák létre stb.), milyen matematikai trükköket alkalmaznak. A fejlesztési rivalizálás tempója és a költségek nyilván a tanítás lehető legnagyobb automatizációjának igényét állítják a fejlesztők elé. A biztonságos és helyes működést azonban nem lehet az emberi tényezők megléte nélkül

[147] Magát a Perplexity rendszert is megkérdeztem ezekről, és a kapott válasz jobb volt, mint amit az általa megadott forrásokat átnézve lehetett volna kapni (tehát több forrást vesz figyelembe, mint amit megjelöl).

biztosítani. Léteznek adat nélküli tanítási modellek, az ún. zero shot modellek, melyekről már esett szó (IV.1.4.). Ám a címkézéshez szükséges segédinformációt ezeknél is az emberi tényező határozza meg [168], tehát ez esetben is igazából a nulladik tényező a meghatározó[148].

Amennyiben megerősítéses módszert szükséges alkalmazni, akkor figyelembe kell venni egy további veszélyt. Korábban (IV.2.2.(3)) példaként említettem, hogy létezhetnek fejlesztői team-beli eltérések, ennek továbbgondolásából adódik, hogy ha az eltérések jelentősek és feloldatlanok (tudattalanok), akkor a rendszert össze is zavarhatják. Egyrészt egy rosszul irányított és tervezett, párhuzamosan végzett megerősítésből jöhet létre ilyen veszély, másrészt a csapaton belüli szemléleti inkoherencia, pl. ellentmondásosan válogatott adatok révén jöhet létre torzítás a végeredményben. Így a team koherens összeállítása és világos irányítása is alapvető tényező a torzítás elleni küzdelemben, de nyilván az operátor szintjén is lehet hibázni akár az adatválogatás, pl. a „megerősítés vagy a büntetés" lépéseiben.

A felhasználók megjelenhetnek egy tanítási felelősség részeként is. Hiszen bizonyos szempontból felelősek saját digitális tevékenységeikért, amelyre fejlett igényfelmerések épülnek. A probléma paradox, hiszen pl. az oldallátogatási statisztikák ezernyi titkos magánéletet elemeznek, sokszor a felhasználók valós életben nem vállalt tartományait. Tehát a rendszer torzít pl. szexualitáscentrikus- tartalmak irányába, a felhasználók egy széles csoportja tiltakozik, holott a figyelembe vett adatok mögött az ő titkos szokásaik is megjelennek. A tudatosan vállalt elvek és a titkoltak az eddigi korokban nem kerültek ilyen tömeges és explicit oppozícióba. Ezért újszerű etikai kérdés jelenik meg abban, hogy hogyan ítélendő meg ez az eltitkolt világ, mely a gépi ítélet mögötti információhalmazt adta, ha az információ létrejöttéért felelős felhasználók sem tartják etikusnak? Itt ráadásul nem egyedi a felelősség, hanem megoszló, akár csak akkor, amikor a felhasználók válaszai alapján lett rasszista az egyik nyelvi modell (ld. IV.2.1.). A felhasználói szint kezelése a tanítási adat szintjén olyan kihívás, amely ellen szinte tehetetlenek a fejlesztők. Akárcsak akkor, ha valaki egy MI-től kapott eredményt épp úgy félreért, ahogyan az emberek értik félre egymást. Továbbá az is nehezen kezelhető, ha az MI egy tudományosan alátámasztott válasza sérti valaki érzéseit. Ezt a legvégső szintet tehát az MI-k fejlesztői nem képesek kezelni, de talán nem is az ő feladatuk.

IV.2.7. További törekvések a torzítás ellen és védelmi vonatkozások

Számos ajánlás fogalmazódott már meg arra, hogy a torzításokat egy adott tűréshatáron belül lehessen tartani. Mivel a szakirodalom és a jelenlegi elvárások általában nem technikai, hanem

[148] Emlékeztetőül: a zebrát ismeri fel az ilyen rendszer, ha definiálják számára, hogy az egy „csíkos ló".

emberi jogi irányból közelítik a kérdést, ezért az alábbi bekezdésekben gyakran használom a torzítás helyett az előítélet szót, a forrásoknak megfelelően.

Az ajánlások lényege általában, hogy fejlesztéskor nagy figyelemmel kell lenni a helyes súlyozás és a tisztességtelen mintavétel elkerülése mellett az elfogultságot kiváltó olyan okokra, mint a velünk született vagy tanult előítéletek. Emellett életcikluson át működő eljárásrendekre van szükség, mely a fejlesztés minden lépésében specifikusan szűri az algoritmikus és adatelőítéletességeket.[149] Amikor végképp elkerülhetetlen a használt adatbázis előítéletessége, akkor annak használóit kell kiképezni a létrejött, elfogult MI helyes alkalmazására. Erre jó példa egy gyermekjóléti prediktív elemző eszköz koncepciója.[150]

Kiemelendő az a kutatási irány, mely a generatív intelligenciát (II.2.3.) hívja segítségül az előítéletesség csökkentése érdekében.[151] Mert bár maga a generatív MI sem mentes még az előítéletektől [172], mégis alkalmas lehet egy kiegyensúlyozatlan adathalmazt egyenletesebbé tenni. Egy egyszerű példával: nem szükséges sok ezer 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). Megjegyzendő, hogy ez épp a különböző kultúrák különbségei – pl. eltérő gesztusjelrendszere miatt – világszinten nem működik jól.

Az is megállapítható, hogy sok kutató szkeptikus a tekintetben, hogy kiküszöbölhető-e az előítéletesség. Egy tanulmány szerint pl. toborzási téren nem megoldott ennek elkerülése.[173, pp. 89] Ám ez a forrás csak pár konkrét emberi okot jelöl meg, ami miatt az objektivitás (szerinte) nem valósulhat meg. Ezáltal csupán gyakorlati példával támasztja alá fenti gondolataimat, melyeket a humánvirtualizáció, illetve a „nulladik tényező" (IV.2.2.) leírásában egy objektívebb és általánosabb perspektívában vázoltam fel.

Összegzésként kimondható, hogy ezért is lenne nagyon fontos sokkal világosabban elkülöníteni egy MI-rendszer előítéletességét és torzítását, mivel véleményem szerint az utóbbi alapos tervezéssel és teszteléssel jól kiküszöbölhető, ám az előbbi, az emberi tényezők miatt, elvileg nem kerülhető el. Egyrészt azért, mert pl. objektív adatok implikálják az „előítéleteket", tehát a rendszer eredményei pusztán „nem mutatnak jól" (de egy hagyományos adatbázis lekérdezése sem mutatna jól). Másrészt a rendszer és az adatok előállítóinak emberi mivolta miatt, hiszen fentebb minden tényezőnél kimutattam a humán hatás jelenlétét.

[149] Ezt alkalmazza pl. a Prevision rendszer fejlesztése a nulladik, kezdeményező fázisától, a tervezésen, a végrehajtáson, a megfigyelésen át a lezáró fázisáig [169].

[150] A témáról többet nyújt a Rhema Vaithianathan vezette kutatók publikációinak gyűjtőoldala [170].

[151] A generatív MI elfogultság elleni használatáról gyakorlati példákkal ld. [171].

Végül egy védelmi vonatkozás. Fontos kiemelni, hogy a torzítások kezelése egyben a vele való visszaélési lehetőségeket is kézzelfoghatóvá teszi. Először hadd világítsak rá erre a tudományos kutatás területéről vett példa alapján. Technikailag nehéz megteremteni annak hátterét, hogy pl. egy magyar tudományos eredményt is relevánsnak fogadjon el egy amerikai automatika. Sőt, például lehet-e a történelmi Magyarország területének történelmével kapcsolatban csak ilyen tényezők alapján ítélni meg a történészek megállapításait? Ha esetleg a környező országok kutatóinak véleménye lényegesen jobban reprezentált a digitális tér releváns részében, akkor egy prompt az ő véleményüket fogja elsődlegesen kiemelni. Ez a példa is mutatja, hogy nem csupán az elfogultság kerülése, hanem az ennek nevében végzett válogatás is védelmi tényező. Hosszabb távon esetleg olyan nemzetközi elvárásrendszer lehet erre megoldás, mely egy ilyen rendszernek annak függvényében ad megfelelő minősítést, hogy ha egy válaszban kérésre sem lenne hajlandó az MI eltekinteni minden érintett álláspontjának vázolásától, még olyan kényes kérdésekben sem, mint a török-örmény konfliktus.

IV.3. HIBÁZÁS AUTOMATIKÁKBAN ÉS „AUTONÓMIÁKBAN”

Alább megvizsgálom, hogy a hibázás kézbentartásának elvi vagy gyakorlati akadályai vannak-e a kétféle rendszerben. Először is visszautalok az előző vizsgálat azon részére, ahol az algoritmizálás és a hardver szintjével (IV.2.4.) kapcsolatban megállapítottam, hogy a hagyományos programozást érintő hibalehetőségek nyilván a neurális rendszereket is érintik. Ez a terület több mint nyolcvan éve fejlődik, ezért a hibatípusok különféle osztályzási típusai és kezelési módjaik nem csupán kiforrottak, de a programozási keretrendszerek számos funkcióval segítik elkerülésüket. Ez a kiforrottság abból is adódik, hogy csak akkor tudtak egy-egy problémát a számítógépekkel megoldatni, amikor készen állt erre a megfelelő formális modell, vagyis a világ bizonyos szegmensét sikerült matematikailag megragadható módon kezelni.

Ezzel szemben a neurális hálók lehetővé tették, hogy a rendszer úgy utánozza a valóságot, hogy annak leképezését nem hozzuk létre.

Ez a működés egyrészt azokat a problémákat vonzotta a technológia felé, melyek matematikai modellel eddig megoldatlanok voltak. Hiszen képes jó eredményt adni az MI-rendszer akkor is, ha nincs pontos matematikai modell, de elegendő adattal van tanítva. Ez igazából nem annyira boszorkányos jelenség: az elmúlt több tízezer évben az átlagemberek is matematikai modellek nélkül tanultak, sőt így oldottak meg sok mindent és alkottak maradandóakat is (itt nem pl. az építményekre gondolok). De jól szemlélteti ezt az érzelmi intelligencia is, ahol nem volt szükség a lelki folyamatok matematikai modelljét felállítani, elegendő volt ezek testi affektusait modellezni, és ezt a modellt megtanítani a gépnek (I.4.). Másrészt bizonyos formalizálható,

elvileg leprogramozható feladatokra is lehet gyorsabb egy MI betanítása, mint egy kód megtervezése, kifejlesztése és tesztelése, – és ez okozza a munkaerőpiaci félelmeket.

Itt térjünk vissza az alapkérdéshez, hogy mennyivel rosszabb („biztonságtalanabb", veszélyesebb) egy ilyen modellnek betanított valóságrészlet, mint egy segédmunkás betanítása, vagy egy hagyományos program elkészítése. Az emberrel összevetve egyértelmű, hogy a monoton munkák, a segédmunkák, sőt bizonyos szakmunkák is elvileg kiválthatóak ezekkel a betanított fizikai vagy szoftverrobotokkal. Mert a tanításon kívül hagyományosan beléjük programozott szabályokat (pl. balesetvédelmet, ügykezelési rendet) a gépek mindig maradéktalanul betartanák. Az emberek fáradnak, lelki válságaik vannak, felvágnak egymás előtt, vagy néha csak nincs kedvük betartani a szabályokat. Tehát az emberekhez képest az egyszerű feladatok ellátásához biztonságosabb rendszerek építhetőek MI-modulokkal, ha a legfontosabb szabályaikat hagyományosan programozott rendszerek adják (más kérdés, hogy ilyen – az emberi fizetések mértékéből kiindulva – még jó ideig nem térül meg. Az átállás csak ott várható ilyen fizikai robotok terjedésére, ahol nagy a tőke és másképpen megoldhatatlan a munkaerőhiány.

Nehezebb kérdés, hogy egy hagyományosan programozott fizikai vagy szoftveres robot biztonságossága tényleg annyival jobb-e, mint az MI-é. Gyakorlati szempontból jelen állás szerint, – a fentebb említett kiforrottságuk miatt, – a hagyományos algoritmusok valóban biztonságosabbak. Az én meglátásom szerint viszont elvileg nem annyival biztonságosabbak, mint amennyire az emberek ezt így érzik. Az eddig ismertetett kutatások alapján felvetődik, hogy a probléma valójában nem a mély neurális hálózat miatti bizonytalanságban van, hanem a rendszerek oly mértékű komplexitásában, melyet egyetlen ember már régen nem képes áttekinteni. Igaz, ez a szintű komplexitás történetileg szorosan kapcsolódik a mélytanuláshoz, mivel a nagyon korszerű rendszerek hatékonyan tudják az ilyen képességeket integrálni. Meglátásom, hogy a kétféle rendszer átláthatósága inkább egymás felé fog konvergálni, ugyanis a rendszerek méretének és komplexitásának növekedésével a hagyományos módszerek átláthatósága romlik, viszont az MI-rendszerek fejlődésével azok egyre átláthatóbbak lesznek.

Nézzük először a hagyományos rendszereket. Ezekben az egyre összetettebb rendszerekben az előző alfejezetben bemutatott minden torzítási hiba megjelenik, a tanítási hiba kivételével. Az adathibákkal kapcsolatban elég megnézni, hogy mekkora a kereslet az adatelemzőkre – holott az adatokhoz elvileg a gépek értenek jobban. Azonban még a strukturált, vagy automatikusan gyűjtött adatok utófeldolgozása is emberi képességeket igényel, a nem strukturált adatok

esetén pedig a jelenlegi MI-k is csupán segédei lehetnek az emberi adatelemzőknek. Az algoritmikus szinten a hagyományos algoritmizálás hibakezelése fejlett. A szintaktikai hibákat[152] a keretrendszer kiszűri, a szemantikai hibák[153] a teammunka és a tesztelések során kezelhetőek. Az igazi problémát a logikai hibák jelentik – akár csak az emberi kommunikáció vagy feladattervezés során. Az ilyen tévedések, az úgynevezett „kognitív torzítás" miatt jöhetnek létre, ennek főbb típusai a következők [174]:

- **Horgonyzási torzítás** (*Anchoring Bias*): a programozók túlzottan ragaszkodnak első implementációjukhoz, „lehorgonyoznak" annál.
- **Megerősítési torzítás** (*Confirmation Bias*): főleg saját fejlesztés ellenőrzésénél a fejlesztő hajlamos olyan teszteseteket írni, amelyek alátámasztják, hogy a kódja helyesen működik, így rejtve maradnak a hibák. Ez nem feltétlenül szándékos, ezért kell független tesztelő vagy korrektor mindenfajta alkotás ellenőrzésénél.
- **Túlzott magabiztosság** (*Overconfidence Bias*): A programozó, aki túlbecsüli képességeit és alábecsüli a feladatok komplexitását.
- **Optimizmus torzítás** (*Optimism Bias*): a fejlesztők hajlamosak túl optimista becsléseket adni a program működésével kapcsolatban. Ez a felhasználás tervezésénél jelenthet gondot (pl. a projektmenedzsmentben okoz jelentős problémákat az ilyen időbecslés).

Ez a néhány példa is jól mutatja, hogy a humánvirtualizációs hibák alatt elemzett problémák csak részben oldhatóak fel az ott említett megfelelő tudásreprezentációk és ontológiák segítségével (V.1.1.) – ezek a klasszikus programozáshoz köthető hibatípusok is emberi (közösségi, önismereti stb.) kezelést igényelnek. És minél összetettebb a rendszer, annál nagyobb az esély arra is, hogy bár a részei jók, az egész összekapcsolásával létrejövő rendszer azonban logikailag már nem lesz kompakt (vö. „viszony szint" IV.2.4.). A nagy rendszerekhez is folyamatosan jelennek meg a javítókódok, mivel[154] a túl nagy komplexitás miatt nem tökéletesek. A nagy szoftverek és operációs rendszerek sérülékenységi mintáit ember már nem képes tesztelni, MI-vel kerestetik ezeket.[175] Ezek is arra utalnak, hogy a rendszerek nagy komplexitása jelenti az elvi problémát, nem pedig a neuronhálós módszerek bevonása. A már elért komplexitást ember nem képes áttekinteni (már régen), ezért a hagyományos rendszereket épp úgy tesztelések által lehet csak egy elfogadható hibahatáron belül működőnek mondani, akárcsak a mélytanuló rendszereket – térjünk is rá ezekre.

[152] „nyelvtani" hibák: az adott programozási nyelv szigorú szabályainak be nem tartása a probléma

[153] „nyelvhelyességi" hibák, melyeket „félreért" a gép, tehát nem a várt eredményt adja

[154] Természetesen azért is, mert a világ változását csak így képesek követni, de most nem ezen van a hangsúly.

A neurális MI tesztelési kritériumrendszere még valóban nem elég kiforrott. A téma jelenlegi fejlődési üteme azonban bíztató, validálható hibakezelési és tesztelési módszertanok kidolgozása nem tűnik elérhetetlennek [176], sőt túl távolinak sem. A legnagyobb tesztelési kihívás, a megmagyarázhatóság még számos kihívásnak néz elébe[155], de a tesztelés a folyamatos fejlődése mellett[156] kiderül, hogy ezt segítheti, ha magukat az MI-rendszereket látják el olyan képességekkel, amely által az eleve átláthatóbb lesz.[178] Tehát a fejlesztések karöltve haladnak a tesztelhetőség növekedésével. Emellett az is felmerül, hogy az átláthatóság hagyományos megközelítéséhez való ragaszkodás helyett ezen a téren is paradigmaváltás szükséges. Például csak arra koncentráljon a tesztelés és az auditálás, hogy egy szigorú hibahatáron belülre kerüljön a rendszer.

Véleményem szerint a hagyományos „magas rendelkezésre állású rendszereknél" (HAS, III.4.3.) napjainkra kialakított alaposság és körültekintés adaptálható lesz ilyen szintet célzó MI-modulok fejlesztésénél is. Tehát hibrid (hagyományos és MI) rendszerek esetén sem lehetetlen ennek a fontos biztonsági szintnek a megvalósítása. Mert bár jelenleg az MI-vel kapcsolatban számos kihívás áll fenn, azonban az ilyen rendszerek nem a bőrszín vagy nemi identitás tényezőit vizsgálnák, hanem pl. veszélyes folyamatok előjeleinek mintázatait keresnék, melyek objektivizálhatók, tehát a gépnek biztonságosan átadható feladatok.

Összefoglalva: mind a hagyományos, mind a neurális MI számára elvileg megoldhatatlan a humán probléma, hiszen egymás felé konvergál az óriásivá növekvő és bonyolódó hagyományos rendszerek átláthatósága és az MI átláthatósága. Az átláthatóság megoldottsága terén jóval előbbre van a hagyományos számítástechnika. Egyelőre. Ám a konvergencia miatt egy idő múlva a vékony MI rendszerek biztonsága nem fog lényegesen eltérni a robusztus hagyományos rendszerekétől. (Egyszerű rendszerek esetében marad a nagy eltérés.)

Tovább görgetve a gondolatot ebből az is következik, hogy az a félelemmel teli szemlélet, mellyel ma az MI-t megközelítik, nem a tudományos megalapozottságon, hanem az irodalmi, film és bulvársajtó felől jövő *benyomásokon* alapul. Figyelembe véve az autonómiával kapcsolatos fentebbi meglátásokat kimondható, hogy nem csupán az MI tudatra ébredése nem fenyeget, de a jelenlegi vékony MI-megoldások „magukról" nem képesek az emberi szándékkal ellentétesen sem tevékenykedni. Csak olyan esetekben lehet alkalmas erre, amikor az alkotók akarata ezt kívánja – csakhogy ez a hagyományos, 50 évvel ezelőtti automatikák esetében is pont így van. Elképzelhető továbbá olyan sorozatos, nagyarányú gondatlanság is, mely miatt a

[155] Ez a dolgozat kimutatja az átláthatóság és magyarázhatóság korlátosságait könyvvizsgáló rendszereknél [177].
[156] Ez a tanulmány 17 módszertant vet egybe [176].

rendszer torzítása óriási, és így akár emberellenes is lehet, – de az ilyen szélsőség az atomerőművek veszélyességéhez hasonlítható, ahol a biztonsági protokollok nem engednek meg hibázást. (Csak azok sorozatos megsértése vezethet balesethez.)

IV.4. A BUKTATÓK TÖRTÉNETI ESETEI ÉS AZOK TANULSÁGAI

Már többször használtam történeti megközelítést, most azonban ezt más hangsúllyal és más irányból teszem. A MI regresszív korszakait és annak tanulságait tekintem át. Az MI-tél (*AI winter*) kifejezést fogom használni, melyet azokra az időszakokra szokás érteni, amikor az MI fejlődése lelassul. Egy-egy technológia fejlődését általában az adott területen folyó fejlesztésekre szánt összegekkel mérik [179], tehát a „tél" egy szókép a kutatási források befagyasztására, a „tavasz" pedig a befektetések megugrásának allegóriája. Jelen elemzés azonban más módon közelít. Először vázolok egy olyan kétdimenziós fejlődési modellt, mely alapján nem mond ellent egymásnak a technológia fejlődése és „stagnálása". (Pontosabban, hogy technológiái áttörés hiánya mellett is hozhat egyre nagyobb hasznot, tehát „fejlődhet".) Ez alapján térek rá az MI-tél korszakokra és azok tanulságaira.

IV.4.1. Többdimenziós fejlődés – stagnálás gazdasági növekedéssel

Egy gondolatkísérlettel szeretném felvezetni a meglátás lényegét. Jelen pillanatban, ha varázsütésre az MI minden továbbfejlesztése leállna, és csupán a mai MI-modellekre tudnánk alapozni: akkor is számtalan új termék és szolgáltatás bontakozhatna ki évtizedeken át. Ezek a meglévő modelleket adaptálnák újabb és újabb feladatokhoz, esetleg csiszolnának 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 a 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 feladatkörét, 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 (AGI[157], vagyis embertudású MI) -rendszer létrehozása, vagy az ember-gép egyesülés könnyűvé tétele lenne. Ezekre sok „prófécia” létezik, megjelenésük realitása azonban véleményem szerint megkérdőjelezhető. Ezért én nem számítok vertikális lépésre az MI-ben a belátható jövőben.

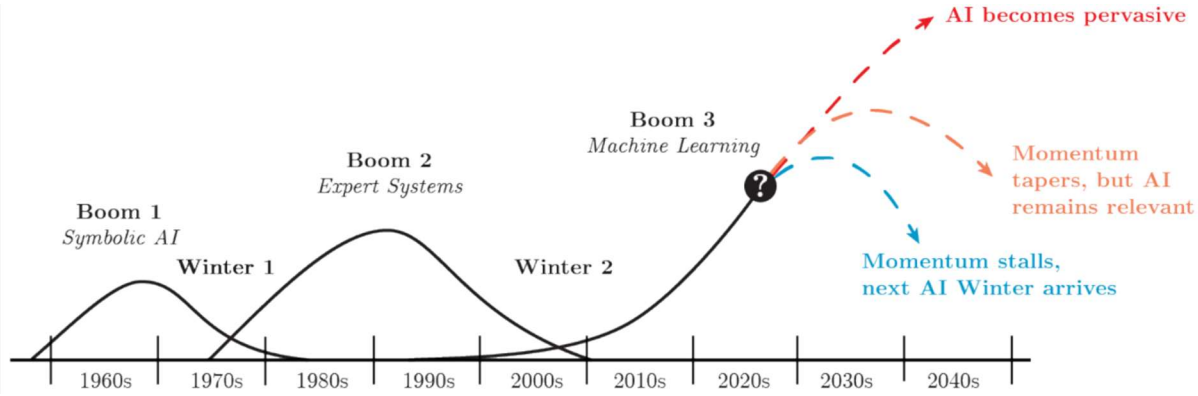
IV.4.2. Az MI telei és tavaszai

A jelenleg tapasztalható MI-fejlődés folytatásának lehetséges forgatókönyve többféle lehet, amint ezt mutatja az 13. sz. ábrán (forrás [180, pp. 2][158]) 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 lépcsőfok elérését prognosztizálja (pl. [27]). Más vélemények szerint a jelenlegi fejlődési tempó ilyen egyszerű meghosszabbítása nem ad megfelelő prognózist: inkább valamilyen mértékű visszaesésre lehet számítani. Jelen tanulmány ez utóbbi megközelítésekre fókuszál. (Az ábrán látható két hullámvölgyről a következő kisfejezet szól.)

[157] Artifitial General Intelligence, tehát pontosabb, de magyartalanabb kifejezéssel Mesterséges Általános Intelligencia.

[158] A tanulmányt a Mitre Corporation, egy nonprofit kormányzati és katonai kutató cég kutatói írták. Visszatérek rá IV.5.3.-ban.

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 (IV.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 még inkább a különbségekre az akkori és a mostani kor között, melyre több következtetés is épül. A szakirodalomban megjelenő tényezők in-



13. ábra: A MI eddigi két téli időszak és lehetséges folytatások (saját átszerkesztés)

terpretálása, 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 AGI és emberies AGI megkülönböztetése (ld. IV.6.) 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 is hasznosítható, míg a direkt és adaptált katonai MI megkülönböztetése a hadtudományi kutatások leírásait pontosíthatja.

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).[159] 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

[159] 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 érdemeik miatt a forrást nem vetettem el.[181]

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 sok tényezőt csak vázlatosan említhetek, és ugyanezért például a lassulás során komoly tényezőként felmerülő jogi területre sem térhetek ki.

IV.4.3. 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 vázolnom, 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 a Toosi és kutatótársai által összeállított esszét választottam ki.[160]

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.[181, pp. 6])

A fősodor befagyása azonban nem jelenti azt, hogy közben semmi nem történik: így a „hó alatt ébredő magok" módjára búvt 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 molekulaszerkezetre következtetni képes rendszer a Stanford Egyetemen [183]: 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 a fejlesztők részére, így sok ágazatot fellendített a szoftverfejlesztésről a robotiparig. Meg kell azonban jegyezni, hogy több kutató a szakértői rendszereket nem sorolja a mesterséges intelligenciákhoz, mivel kívülről nem

[160] A részletesebben érdeklődők számára ajánlható tanulmány, mely sok adatot és nevet is tartalmaz.[182, pp. 11]

tanulnak, inkább szabályalapúak.[161] 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 egyetért, hogy az 1990-es években a szakértői rendszerek felívelése abbamaradt. 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.[181, pp. 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.[162]

De „a hó alatt” végig zajlott a fejlődés. Már 1980 körül (a második tavasz alatt) elkészültek az első konvolúciós neuronhálók.[186] 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 kivirágozzon.

IV.4.4. 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 hátterében is. Általános jellemzője ezeknek a technológiai teleknek, hogy két főszereplőjük van: (1.) *a kutatók*, akik a túl optimista előrejelzéseket adják; (2.) a *döntéshozók*, akikben a túlzott remények keletkeztek, ezért támogatásról döntöttek, majd a csalódás miatt nem invesztáltak többet a technológiába.
    b. **Ez nem hype-ciklus.** Fontos, hogy ez a fajta hype-következmény merőben eltér a ma már jól ismert a hype-ciklus [187] típusú prognózistól. Ugyanis, mint a történeti leírásban is rámutattam, a lefagyott kutatások részben fagyottan maradtak, mindig más paradigmák mentén indult be valami új MI-tavasz.

[161] Nem tekinti MI-nek pl. [184].

[162] Ezt a meglátást, a második MI-tél kezdetekor a DARPA egyik intézetének igazgatója fogalmazta meg.[185].

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. A h**ardver probléma több oldala**

   a. **A hardverhiány és az adathiány, mint ok.** A szakértői rendszerek kudarcában szerepet játszott, hogy a hardvergyártók nem tudtak megfelelő alapokat adni hozzá.[182] 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 „elhivés" állapotába, mely nem legitim és tudományos közelítés.

   c. **A jónak hitt modell bizonyítható hiányosságai.** Az első tél egyik kiváltó oka volt, hogy matematikailag is sikerült rávilágítani a megközelítés hibájára, hiszen a Perceptron [70] elméletileg sem képes megvalósítani a XOR függvényt.[188]

   d. **A tanító adatok hiányába vetett hit.** Egy-egy tanító metódus 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, ha szembenéznek azzal, hogy ezek a modellek mennyire leegyszerűsítették a valóságot. Hiszen egy szimplifikáció mindig kompromisszumokat hoz, eltekint dolgoktól a jobb kezelhetőség érdekében – ezáltal alkalmatlan arra, és 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.[162]

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 elhivés. Ez a jelenség szempontunkból is igen tanulságos, főleg ha a három okra is rámutatok, melyre visszavezethető: (1.) az akkori túl bizakodó racionális és progresszivista (a fejlődésben hívő) korszellem; (2.) egy nyájeffektus, 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-tavasz esetében az emberi gondolkodást akarták lemásolni [189], 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.[182, pp. 9] Így a második tél 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.[184]

IV.4.5. A korszakok eltérései

A fenti történeti és okozati vázlat megalapozta, hogy a jelenlegi helyzet és a korábbi két tél közötti fontosabb eltérésekre rávilágítsak. Abból indulok ki, hogy a jelenlegi helyzetet az olvasó ismeri. 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-tél 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 sok ezer 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.

IV.4.6. 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 hátterében), 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 tömegtájékoztatás és a híreket befogadó tömegek. A „mesterséges intelligencia" egy kattintásvadász kifejezés a bulvárújságírásban, mely tömegével ontja az információhamisító cikkeket. De a sajtó többi része sem védett: ezeknél a hamisítás oka sokszor az, hogy sok ezer (más területen talán jártas) szakember képzetlenül fogalmaz meg véleményt az MI-vel kapcsolatosan. Ki is mutatható, hogy MI-publikációkban tendenciózusan kezd összemosódni a spekuláció az igazolható eredményekkel.[190] 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 hypolók, 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 határozott kerülését. Radikális megjelenésében akár az MI korlátozását vagy betiltását is eredményezheti, ha ezáltal valaki szavazatokat szerezhet.[162]
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.[191] Ezeknek jelentős lökést

adott Kurzweil szingularitás elmélete,[163] mely szerint a technológia nemsokára olyan ember és gép egybeépítést 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 a humántudományok (pl. nyelvtudomány, pszichológia) olyan irányú fejlődése is támogatta a haladást, melyek kifejezetten a gépi tanulás alá dolgoztak. Ezek miatt az évtized második feléig sok kutató értett egyet abban, hogy a szingularitás csak idő kérdése [192, pp. 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.

3. **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 kifullad, az újabb mozgalmakban (beat, hippy) az érzelem kap óriási szerepet, majd a kilencvenes években kibontakozik a posztmodernnek nevezett korszak, ahol az irracionalitás visszaszivárog 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.[193] De hasonló az egyes mozgalmak és rajongótáborok világnézete is. Példaként lehet hozni erre a Turing Church mozgalmat,[164] 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.

IV.5. A HORIZONTÁLIS ÉS VERTIKÁLIS TANULÁS VÉDELMI KIHÍVÁSAI

A legtöbb MI-vel foglalkozó munka, és az azt népszerűsítő propaganda evidenciának veszi, hogy az embereknek „csak” meg kell tanulniuk az MI-t jól használni, mint ahogyan megtanulnak egy idegen nyelvet. Nem találtam elemzést az ezt akadályozó tényezőkről, melyek az MI,

[163] Ez 2006-ban jelent meg, de mára magyarul is elérhető: [29].

[164] Ennek neve a szójáték a Chuch-Turing matematikai tézissel, ezért nem fordítottuk le. Rövid összefoglaló a vezetőjétől: [194].

vagy más, hasonlóan nagy lépést képviselő diszrupciók kiváltanak az emberekben. Ebben az alfejezetben saját meglátásaimat vázolom, és bemutatom a téma védelmi aspektusait.

IV.5.1. A többdimenziós tanulási modell

Egy technológia fejlesztésében sok az objektív tényező, ezért jóval kiszámíthatóbb maga a fejlesztés, mint az eredmény fogadtatása.[195] 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 tanulás megkülönböztetésére épül.[196] Ezek lazán kapcsolódnak a fejlődés fentebb említett horizontális és vertikális megkülönböztetéséhez. Röviden a *horizontális tanulás* a meglévő szemléletet csiszolja, egészíti ki további információkkal, míg *vertikális tanulás*nak azt nevezem, 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, a 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 tudja azt. Ennek oka lehet az, hogy a befogadó számára érdektelen a téma, de az is, hogy az a korábbi megértés, mely

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

Más hasonlattal: a horizontális tanulás néha olyan, mintha különböző méretű legókockákat akarnánk egymásba pattintani, s mikor ez nem sikerül, az inkompatibilis legódarabokat csak beszórjuk 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 szemléletkialakító vagy szemléletváltó tanulásnak: ezáltal képes az alany kihasználni az új paradigma erejét, hiszen egy új hidat 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 „rááll” 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 rakják, amelyek 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 életük 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. Legtöbben 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 el kell telni. (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 pontokba 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 polgáraink számára. Pontosabban jelenleg nem tudjuk elképzelni, hogy ezt el lehessen fogadtatni. (Ezt a gondolati szálat védelmi aspektusból majd folytatom, ld. VI.4).
- 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 jelenség még problémásabb a döntéshozók esetében. Mert ha ők benne rekednek egy régi szemléletben, az (döntési pozíciójuk miatt) sokkal erőteljesebben visszafogó hatású 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 kedvük 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 szeretik a bejáratott dolgokat az életben, nem szeretnek sokszor nagyot változtatni, főleg nem önmagukon.
- Minél erősebb kognitív képességekkel rendelkezik az MI, annál erőteljesebben érvényesül majd a gondolkodásmód leváltásának fontossága a megfelelő ember-gép együttműködésben. Másképp fogalmazva minél nagyobbat fejlődik az MI, annál nagyobb vertikális ugrásra lesz szükség. (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 lehetőségeit 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 ódzkodik tőle, hiszen valami új, valami programozásszerű. 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 önváltoztatásra (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. – ez esetben viszont épp a kényelem ellenében szükséges haladni. Lehet erőltetni a kötelező gyorstalpaló képzéseken való részvételt, és ezáltal kimutatható lesz, hogy többen használják az új lehetőséget. Ám ez nem fogja kialakítani bennük azt az új szemlélet, 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 rendeletileg 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 ölteni, 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 szemléletváltásra 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.

IV.5.2. IQ-fenntarthatóság (a tehetségek végessége)

Mai kultúráink 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 ún. „értelmiségi" 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 szakmunkában feladatot végezni, ezek terjedése azonban a várhatóan magas áruk (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 kérdésről folynak kutatások (pl. [197]). Viszont kevésbé kutatott a másik terület – amely szerintem sokkal nagyobb gond, – hogy nem is születik elegendő olyan képességű ember, akikből ez az óriási tömegű magas szakmaiságú munkaerő kiképezhető lenne. Tehát azért lehetetlen elegendő jó orvost, jó pedagógust, jó mérnököt, jó kőművest találni, mert nem létezik ennyi tehetség világszinten. Természetesen az un. 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 a kifejezés igazából pontatlan. Ugyanis a fentiekhez hozzáadódik, hogy a magas szakmai tudás mellé egyre több egyéb adottság vagy nevelt készség tár-

sulása is szükséges: a stressztűrés, a magas koncentrációs készség, a jó kommunikációs készség, a kedvesség, a 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 tehetségkomplexum-fenntarthatóság" – ám ez igen nehézkes elnevezés, maradjunk a címbelinél.

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

IV.5.3. A várható MI-nyár és ennek kiaknázhatósága

MI-nyár kifejezés alatt, egy lényeges technikai áttörés nélküli technológiaterjedést értek. Ebben a nyárjellegű fejlődésben a változás lassul, de a felhasználások száma növekszik. A jelenlegi helyzet a II. fejezetben vázolt számos irányú fejlesztés alapján véleményem szerint így értékelhető. Vagyis – a fejezet elején (IV.4.1.) definiált kifejezésekkel élve, – a nagy sebességgel zajló horizontális fejlődés folytatódása várható az MI-ben, de vertikális progresszió nincs kilátásban. Más szóval a jövőben, – számos kisebb ötlet és innováció bevezetésével – tovább tökéletesednek a most ismert MI-képességek, de lényeges lépésre, igazi általános MI létrejöttére véleményem szerint belátható időn belül nem lehet számítani. A terület és az ilyen irányú befektetések jelentős bevételekre számíthatnak.

Pillantsunk rá ennek védelmi oldalára. A fentebb idézett (IV.4.2. eleje) Mitre-elemzés [180] kiindulópontjában az áll, hogy ha az MI fejlődési sebessége lassulna, az jelentős és hátrányos hatással lenne a védelmi rendszerekre. Az a megközelítés azonban a korábbi MI-telekhez hasonló korszaktól szeretné megóvni a társadalmat. Ezzel szemben fentebb én egy teljesen más jellegű, „gyorsulással egybekötött lassulást" vázoltam, amely ilyen fajta visszavetést nem okozna. Sőt, ez a nyár kiaknázható védelmi téren. Rá is térek arra, hogy hogyan lehet egy ilyen forgatókönyv esetén a hátrányunkat előnnyé 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 világelső fejlesztések lényegi változást nem hoznak. Így aki megtanul például jól promptolni, az öt év múlva is képes lesz használni ezt a tudását. Az nem probléma, ha a terjedés közben tovább halad. Ugyanis az említett vertikális tanulási jelenség (IV.5.1.) problémája miatt min-

denhol időbe telik ennek az új paradigmának elégséges számú vezető szemléletébe való beépülése, és a dolgozóknál a felhasználói szintű használat elterjesztése. Viszont az ingyenes szolgáltatásokon és a nyílt forráskódú modellekből átdolgozott rendszereken is hatékonyan végezhetőek ilyen szakmai képzések (és önképzések). Jelenleg is jobbára ez történik, csak kissé lomha tempóban – így a lemaradás nem csökken. Ám ha a MI-technológia fejlődési sebessége lelassul, és nem jön el a „nagy áttörés", akkor érdemes lenne a jelenleginél nagyobb képzési sebességbe kapcsolva megpróbálni minél többet ledolgozni a hátrányból a humánerő terén. Ezáltal létrejöhetne egy olyan apparátus, amely a digitális szuverenitást biztosító 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. Sajnos a hazai mentalitáshoz közelebb áll az, hogy nem képzi magát az, aki azt látja, hogy önváltoztatás nélkül is megmarad a munkája. Pedig a közeljövőben reálisan megvalósítható MI-szolgáltatások egyre több sematikus „értelmiségi" munka kiváltását fogják lehetővé tenni még akkor is, ha a fejlődés lelassul (ha nem lassul le, akkor gyorsabban). Ilyen megoldások bevezetésében a nagy irodákkal rendelkező magánvállalatok fognak az élen járni, a kreatívabb dolgozókat átképzik, a többieket elbocsájtják. Az állami szféra vagy a kisebb vállalatok várhatóan még évekig embereket fognak alkalmazni automatizálható feladatokra, akik ezalatt nem készülnek a váltásra mondván, hogy úgysem lesz pénz az MI bevezetésére. A lakosság felkészítése nélkül a szavazók nem fogják tudomásul venni, hogy az ország versenyképessége függhet ettől, így szinte bizonyos, hogy az erre szánható összegeket inkább az örök problémát jelentő szociális szférák valamelyikébe kívánják majd átcsoportosíttatni, akár tüntetéseket is szervezve ezért. Ilyen megnyilvánulások megjelenésére hazánkban véleményem szerint 4-5 éven belül lehet számítani, később pedig a kérdés akár a választások egyik kulcskérdésévé is válhat. Tehát ha nem sikerül megfelelő motivációkat találni és a fent leírt módon, elegendő vezető, kolléga és dolgozó állampolgár fejében a szemléletváltáshoz elengedhetetlen szintet „meglépetni" (vagyis rávenni őket, hogy lépjék meg – hiszen helyettük más erre nem képes), akkor mindkét forgatókönyv szerint hátrányba kerülünk. Ha lassul a fejlődés üteme, akkor nem tudjuk kihasználni annak lehetőségét, ha nem lassul, akkor pedig egyértelműen lemaradunk.

Ezzel a gondolatsorral azt is bemutattam, hogy a technológiai és humán tényező egymás hatását mindig erősíti: vagy a progresszív hatást vagy a regresszívet. Ha sikerül egy új dolgot divatossá tenni, akkor kis erőkifejtéssel terjeszthető, ha pedig „ciki", akkor nagy erőbevetéssel is nehezen terjed egy technológia vagy brand. Mind a felfelé vivő spirál, mind pedig a lefelé mutató örvény öngerjesztő folyamatként működik. Az előbbi ráadásul a befektetést is serkenti,

az utóbbinál pedig sokkal kisebb lesz a technológiai-befektetői motiváció is saját MI-rendszereink fejlesztésére és élvonalba emelésére. Így az emberek technológiakövetése felívelést hoz, a humán lemaradás pedig 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 leszünk.

IV.6. RÉSZÖSSZEFOGLALÁS: AZ MI BIZTONSÁGI KIHÍVÁSAI

Összegzés. Az MI-vel kapcsolatos kihívások áttekintése során a védelmi célhoz (C2) számos hasznos meglátásra jutottam. Kevesebb konkrét adalékokat kaptam a terminológiai célhoz (C1), azonban egy új fogalomhoz szükséges a megfelelő szemlélet kialakítása is, ehhez járult hozzá ez az elemzés. Az itt bemutatott tényezők áttekintése rávilágított arra is, hogy az MI újszerűsége milyen védelmi kihívásokat generál, ehhez hasznos volt áttekinteni és osztályozni a neurális hálók miatt felmerülő problémák típusait. A védelmi problémához (P2) kiemelt fontosságú az MI hibázásának elemzése is, ezt elvégezve az eredményt összevetettem a hagyományos automatika tévedéseivel. A körképhez két humán témakör tárgyalására is szükség volt: egyrészt történelmi tanulságok levonására, másrészt az MI tanulhatóságának vizsgálatára.

P1-gyel és P2-vel egyaránt kapcsolatos következtetések

R4.1.: Léteznek olyan elvileg nem leküzdhető problémák, melyek egyaránt megjelennek az általunk alkotott hagyományos gépekben és az MI torzítások mögött (IV.2-3.).

- R4.1-A: Az emberi kogníció és a világ leképezési problémái (végessége, hibái, szubjektivitásai).
- R4.1-B: A rendszerek robusztus komplexitása (ember általi áttekinthetetlenségük).

R4.2.: Az óriásivá növekvő és bonyolódó hagyományos rendszerek és az MI átláthatósága egymás felé konvergál (IV.4.).

P1.gyel kapcsolatos következtetés

R4.3. A R4.1-A-val összefüggésben (III.5 és IV.2.2. alapján) a leképezés és a kogníció kifejezések megjelenítése javasolt a technológia meghatározásában, mivel említésük utal a technológia elvi határaira.

P2-vel kapcsolatos következtetések

R4.4. A jelenlegi MI újdonságai egyben a jelenlegi fejlesztések fő problémáit is jelentik, melyek feloldása nem garantált, de bíztató (IV.1.).

R4.5. Az R4.1&2. folyományaként levonható, hogy a vékony MIrendszerek biztonsága elvi szinten hasonló, mint a hagyományos számítástechnikáé (az utóbbi azonban gyakorlati megoldottság terén még jóval előbbre van) (IV.2-3.).

R4.6. Az MI-fejlesztésekben a kulturális és a társadalmi különbségek meg fognak jelenni (IV.2.2.(3)).

R4.7. Gyakorlati (egy jelenlegi fejlettségi) szinten nem leküzdött problémák, melyek mindkét rendszertípus esetén ugyanúgy, csak teszteléssel oldhatóak meg (IV.2-3.):

R4.7-A: A hagyományos rendszereknél a robusztusság és a rendszerrészek viszonya;

R4.7-B: MI esetében az átláthatóság és a feketedoboz-jelleg egyéb következményei.

R4.8. Az R4.2 és R4.7. miatt a rendszerek biztonsági megközelítéseiben is elképzelhető egy paradigmaváltás szükségessége (IV.2-3.).

R4.9. Az MI-től való félelmek gyakran túlzóak, melyet a bulvársajtó rendszeres tudománytalansága is gerjeszt (IV.3.).

R4.9-A A jelenlegi vékony MI csak az alkotók akarata, vagy sorozatos nagyarányú gondatlansága miatt lehet képes az emberi szándékkal ellentétesen tevékenykedni, és nem képes „csak úgy” önállósulni, netán tudatra ébredni.

R4.9-B Biztonsági szempontból az atomerőmű-technológiához lehet hasonlítani.

R4.10. Az R4.1, R4.5, R4.7 és R3.9 alapján, az MI autonómia biztonságtalanságához kapcsolódó általános vélemény a jövőben akkor is hátráltatni fogja az MI-be vetett bizalmat, ha a két rendszer biztonsága között statisztikai szempontból nem, vagy alig lesz eltérés.

R4.11. A korábbi két MI-tél tanulságait korunkban is figyelembe kell venni (IV.4.).

R4.12 Szükséges megkülönböztetni horizontális és vertikális fejlődést, vagyis a lényegi technológiai lépéseket és az azokon belüli fejlesztéseket (IV.4.1.).

R4.13. Az emberek vertikális tanulásának megvalósulása az MI (és az egyéb új paradigmák) hatékony használatához elengedhetetlen feltétel, de ennek személyes megvalósítása rendkívül nehéz (a hozzá szükséges képességek és lelki energia miatt) (IV.5.1.).

R4.14. Az R4.11-13 alapján az MI hosszú nyara várható (vagyis az MI fejlődési sebessége lassulhat, de erőteljes terjedése van kilátásban) (IV.4.).

R4.15. Az R4.14 szerinti hosszú nyár kiaknázható és kihasználandó lenne a lakosság MI ismereteinek célzott bővítésére, és a digitális lemaradásból való felzárkózásra (IV.5.3.).

V. A MESTERSÉGES INTELLIGENCIA FOGALMÁNAK ÚJRAGONDOLÁSA

Az eddigi kutatások alapján elérkeztem a terminológiai problémakör (P1) vizsgálatának befejezésére. A K5-ben[165] feltett kérdésekre az első négy fejezetben is kaptam bizonyos válaszokat, de a kérdéskör elfogadható megválaszolásához hátra van még néhány részkutatás. Azt kéne majd bizonyítani, hogy *az MI jelenlegi meghatározásaiból fontos aspektusok hiányoznak* (H1). Ezt a hiányt a C1-ben[166] két részre bontottam: egyrészt bizonyos tartalmak bevonása nélkül félreérthető a fogalom, mert kulcskifejezések *maradnak* ki a definíciókból, másrészt a fogalmat következetlenül használják, homályos marad, mivel *beleérthetőek* egymást kizáró jellemzők, tehát az MI kifejezés világosabb lehatárolása is szükséges lenne. Már a korábbi fejezetekben is azonosítottam a definíciókból kimaradó kifejezéseket. Jelen vizsgálatban ezek körét tovább szélesítem, és sor kerül arra is, hogy a fogalom elégtelen lehatárolásának okaira rámutassak. A vizsgálatokat az utóbbival kezdem, rendszerezve azokat az eseteket, amikor az MI kifejezést nem neurális rendszerekre alkalmazzák (V.1.). Ezek alapján prognosztizálom egy jövőben várható pontatlan használatát is (V.2.). Majd az információ és a hozzá kapcsolódó tudományok felől is közelítem az MI-t, hiszen így tárul fel az a széles össztudományi összefüggés, mely valójában a technológia mögött áll (V.3.). Ezek után, összesítve az első öt fejezetben feltárt információkat, le tudom zárni H1 hipotézisem igazolását is (V.4.), bár maga a bizonyítás a dolgozat végén lévő összegzésben található. Végül, az utóbbi alfejezet végén, saját javaslatokat teszek új MI definíciókra, ezáltal a C1 célt is elérem.

V.1. A „DETERMINISZTIKUS MI”, AVAGY HOGYAN HÍVJUK A NEM NEURONHÁLÓVAL MŰKÖDŐ MI-KET?

Determinisztikus MI-nek nevezem, amikor olyan programkódokat is MI-nek hívnak, melyekben nincs neurális háló okozta feketedoboz. Az, hogy ezeket is beleértik a fogalomba, igen ellentmondásossá és zavarossá teszi az indeterminisztikusságból adódó problémákkal (ld.

[165] K5: A terminológia kérdésköre: *Milyen esetekben vetődik fel, hogy hiányos vagy zavaros az MI-fogalom használata, hogyan tehető javaslat az MI-terminológia javítására, és hat-e ez a klasszikustól eltérő adatkezelési paradigma „informatika” kifejezésére?*

[166] C1: Minél több tényezőt azonosítani, melyek nélkül az MI meghatározása félreérthető, vagy amelyeket beleértve a fogalom zavarossá válik, valamint javaslatokat tenni a feltárt problémák feloldására (új fogalmak, felosztások, más kifejezések stb.).

IV.1.) küzdő modern MI-értelmezést. Ez a fogalmi zavar a szabályzás és az elvi biztonság útjában áll, ezért C1 cél miatt muszáj megvizsgálni azt, hogy milyen nem-neurális rendszereket szokás MI-nek hívni. Ám ebből az elemzésből – a biztonsági vonatkozások védelmi fontossága miatt – C2-höz is hasznos adalékokra számíthatunk.

Mint látni fogjuk, a túl tág MI-fogalom hátterében az a történeti-emberi jelenség áll, amelyet szinte mindenki érzékel: hogy a nyelv nem képes követni a technológia ilyen gyors fejlődését. Ezért jött létre, hogy többen, több okból, nem feltétlenül a neurális rendszereket értik MI alatt. Sokan például minden tanuló rendszert MI-nek mondanak, holott egy rendszer hagyományos programkód által is képes tanulni (amit én pszeudo-tanulásnak neveztem II.2.4.), nem csupán mélytanulással. Mások még egyszerűbb szoftvereket is az MI-hez sorolnak, mondván, hogy az ember intelligens vonásainak valamelyikét utánozza. Nem leltem fel olyan kutatást, amely rendszerezné, hogy mikor használják az MI kifejezést nem neurális rendszerekre. Ezért magam azonosítottam azt a három irányt, amely alapján ezt a vizsgálatot lefolytatom:

1. a **kutatók**, főleg a kezdetekben elkezdték használni az intelligencia szót számos, akkor újszerűnek számító informatikai modellre és ötletre;
2. a **felhasználók** sokszor nem tudják megkülönböztetni a mélytanulási rendszerek szolgáltatásait a hagyományos számítástechnikától;
3. a **piaci szereplők** profitvadász fogalomhasználata fokozza az előző két pont hatását.

Az első három rész ezeket mutatja be, majd terminológiai javaslatokat vetek fel a probléma feloldására, végül a három eset kezelhetőségét elemzem. Az alább elemzett jelenség a „*terminológiai degradációs modell*” nem csupán az itt bemutatott esetekre alkalmazható, hanem általánosítható és sok más esetben is használható (elsősorban a technológiában használt kulcsszavak esetében). A következő alfejezetben (V.2.) reprezentálom is a modell alkalmazását: egy fogalom (az AGI) használatának várható torzulását.

V.1.1. Amit a szakemberek egykor intelligensnek hívtak

A számítógépek terjedésével, ahogyan a gépek egyre több „intellektuális” problémát voltak képesek megoldani, jól hangzott, hogy intelligensnek kezdték el nevezni a gépek bizonyos szolgáltatásait. Emiatt a szóhasználat miatt az MI első hulláma (I.2.1.) csak részben tekinthető igazi hullámnak. Bár a neuronháló már létezett, de nem társították az intelligenciautánzáshoz, sőt használata máig sem követelmény az MI-vel kapcsolatban. Arra sem volt konszenzus, hogy milyen programokat lehet intelligensnek hívni. Az intelligens ágens (I.2.4.) kifejezés újabb impulzust adott annak, hogy könnyen besoroljanak valamely technológiát az MI-fogalom alá. Itt

csoportokba rendezve listázom[167] a gyakoribb olyan MI szóhasználatokat, amelyeket eredetileg nem neurális technológiaként soroltak ide. Megjegyzendő, hogy azóta több ilyen problémához találtak MI-alapú megoldást, vagy MI-vel is támogatják a hagyományos modellt, így egyes determinisztikus MI-rendszerek ma már valóban indeterminisztikusak.

1. **GOFAI** (*Good Old-Fashioned Artificial Intelligence*, „jó kis régimódi MI"). Megfelelő automatikát biztosítanak olyan esetekben, amikor a feladat jól formalizálható. Fő előnyük, hogy nincs szükségük sok adatra és nagy számítógépes erőforrásokra, tehát sokkal olcsóbbak. Ugyanerre használják még a *hagyományos MI* kifejezést vagy következőkben felsorolt elnevezéseket is. Néhány rendszertípus, melyet ide sorolok:
    - **Szakértői rendszerek:** Hozzájuk kötődik az MI első nagy elterjedése és a második MI-télnek nevezett időszak (IV.4.2.). Ezek a rendszerek a tudást „ha… - akkor…" típusú szabályok formájában kódolják le, ezért a „**Szabályalapú rendszerek**" elnevezés is ezekre utal. Előnyük, hogy a szabályokkal eleve magyarázatot adnak döntési folyamataikra, ez pedig különösen értékessé tette őket olyan területeken, mint az orvostudomány, a mezőgazdaság és az üzemeltetés. A szakértői rendszerek deduktív következtetést alkalmaznak, vagyis a következtető motor úgy jut eredményre, hogy végigmegy a szabályokon. Tanulni nem képes, kézileg kell a szabályhalmazt bővíteni.
    - **Esetalapú érvelés (Case-based reasoning, CBR)**: Az előző típus továbbfejlesztése [199], itt is alkalmaznak szabályokat, azonban ezek alapján a rendszer induktívan is következtet, vagyis képes hasonló, de új problémákat is megoldani. A hasonlósági számítások lényege, hogy a rendszer először megkeresi a leginkább hasonló korábbi eseteket, majd ezek megoldásait alkalmazza az aktuális problémára. Az ilyen rendszer képes tanulni, de csak a pszeudo-tanulás (II.2.4.) értelmében: csupán eltárolja a feldolgozott eseteket és később ezeket is használja.
    - **Ontológiák** és a tudásreprezentáció bizonyos típusai (pl. szimbolikus MI, IV.2.2.(2)).
    - **Hagyományos viselkedésalapú MI**: Az emberi viselkedés megértésére, előrejelzésére és reagálására összpontosít.[200] A viselkedéstudomány eredményeinek gépi reprezentációján, pontosabban annak formalizálásán alapul, tehát még nem mélytanulással ismeri fel a viselkedést, mint pl. az affektív számítástechnikában bemutatott rendszerek. Ehelyett pl. a felhasználói interakciókból származó adatokat használja fel, ezek alapján

[167] Ehhez a kiindulópontom ez a cikk volt [198], de a listám ettől eltér, és ezen túlmutat.

lesz képes fenyegetések észlelésére és a valós idejű változásokhoz való alkalmazkodásra, vagy a javaslatok személyre szabására. Modularizált, egyszerű viselkedési egységekből épül fel a rendszer, amelyek együtt komplex viselkedést eredményeznek. Más esetekben egy nem túl komplex viselkedés besorolására képesek, ez a megközelítés a viselkedés-output terén, tehát a robotikában, vagy a webportáloknál régóta használatos (a következő alfejezetben erre visszatérek).

2. **Gráfalapú tudásrendszerek**
   - A **tudásreprezentáció** bizonyos típusai, pl. szemantikus hálók (IV.2.2.(2))
   - **Döntési fák:** Olyan fastruktúrákat használnak, ahol a belső csomópontokban döntési szabályok vannak megfogalmazva, a levelei pedig ezek kimeneti esetei. Jól alkalmazható a programozott döntések hatékony megvalósításán túl osztályozási és predikciós feladatokra is.[201]
   - **Bayes-hálók, egyéb valószínűségi modellek**: Itt is gráfstruktúrában ábrázolják a változók közötti kapcsolatokat, de ez alapján valószínűségi következtetéseket hoz a rendszer.
   - **Keresési algoritmusok**: Ezeknél a gráfba rendezett világrészlet-leképezés a gyors keresést teszi optimálisabbá. Ismert alkalmazása az útvonaltervezés.
3. **Kényszerkielégítési problémák** (*Constraint Satisfaction Problem*, CSP): Kombinatorikus problémák reprezentálására és megoldására biztosít formális kereteket. A CSP számítási feladatban a rendszernek úgy kell a változók halmazát értékekkel ellátni, hogy minden megadott korlátozás teljesüljön.[202] Tipikus példái a különféle órarendi vagy ügyeleti beosztások.
4. **Fuzzy rendszerek** (ld. II.3.2.)
5. **Biológiai ihletettségű rendszerek** (II.3.3.)

Ez a néhány példa jól mutatja, hogy a tudományos világban a mélytanulás terjedése nem hozott fogalmi változást, azt a mai napig egy kalap alá veszik a fenti rendszerfejlesztési irányokkal. A fejlesztőknél ez nem okoz fogalmi zavart, mivel szakmai tudás alapján a kontextusból világos, mikor melyikről van szó. Emellett, mint említettem a fenti rendszerek jó része összeolvadóban van a neuronhálós fejlesztésekkel: ezekre a hagyományos platformokon elért eredményekre lehetett alapozni pl. a mélytanuló modellek fogalmait, következtetőképességét, viselkedési, szociális vagy egyéb szolgáltatásait.

V.1.2. A rendszerek összemosódása a felhasználóknál

A rendszerek összemosódása alatt azt értem, hogy a felhasználókban nincs világos tudás arról, hogy miket is használnak, és ezek hogyan működnek. A szolgálatáscentrikus fejlesztéseknek köszönhetően nem is kell tudniuk erről, ami önmagában jó, mert ez teszi lehetővé a fejlesztések könnyű terjedését, viszont most, amikor nagyon eltérő paradigmájú technológiák jelennek meg a háttérben, a biztonságos használat érdekében fontos lenne bizonyos technológiai alapismeret és megkülönböztetés. A felhasználóknak igen nehéz lépést tartani a fejlődéssel, könnyen azt hiheti valaki, hogy nagyjából tudja, mit is használ („nyilván úgy működik, mint régen"), holott nem ez a valóság.

Az ilyen jellegű „hit"-beli tévedések informatikai biztonság szempontjából is problematikus voltára példának hozható az automatikus felhőszinkronizáció megjelenése. Sok tájékozatlan felhasználó nem igazán volt tisztában azzal, hogy mivel jár egy felhő hátterű applikáció. Kezdetben nem tudták az alapértelmezett szinkronizációs automatikát kikapcsolni, így magán, céges és akár érzékeny adatok is kikerültek az adatgazda kontrollja alól. Hasonló az itt tárgyalt tévedéstípus is, amikor a felhasználó keveri a determinisztikus és az indeterminisztikus MI-t, azaz a mélytanulásos neurális és a pszeudo-tanulásos hagyományos számítástechnika szolgáltatásai mosódnak össze.

A szétválasztást nehezíti, hogy a feladatokat gyakran tényleg együtt hajtja végre kétféle rendszer: a neuronhálós és a klasszikusan programozott megoldások, adattó megoldások és hagyományosabb relációs lekérdezések, valamint felhők és helyi applikációk. Ezért sok felhasználónak úgy is tűnhet, hogy minden hasznos és bonyolult szolgáltatásban elbújtattak egy neurális MI-t, akkor is, ha csak egy hagyományos programról van szó. A fő gond az, hogy még az előző technológiát sem értik a legtöbben és máris itt az új. A hatékony használathoz azonban ismerni kellene pl. az adatmezőkben „tanuló" és a neuronhálóban tanuló rendszerek közötti eltérést. A számítástechnika történetére egyébként jellemző, hogy az elavult technológiák nem megszűnnek, hanem más területen használják fel őket. Itt épp tanúi lehetünk ennek a folyamatnak. Az ipar sokkal jobban tiszteli a hagyományait, mint a társadalom, már csak azért is, mivel egy kiforrott áramköri tervet vagy kódot könnyebb újrahasznosítani, mint a fizikai világban kezelni a szemetet.

Az olcsóbb fejlesztés motivációja mellett fontos az is, hogy ezek a rendszerek vegyesen igazán jók. A felhős példánál maradva: az offline funkciók mindig elérhetőek, de online funkciók által sokszorozódik meg a gép tudása. Az adatfeldolgozásban a hagyományos programkódok

és adatbázisok egyelőre sokkal pontosabbak és biztonságosabbak, de az MI által az adatfeldolgozás igazi információfeldolgozássá avanzsálhat (ld. V.3.3.). Alább egy MI és egy hagyományos rendszer összemosódását mutatom be alaposabban, webshop és a szakirodalom-kereső rendszerek példáin keresztül. Ez szemlélteti, hogy miért képesek ezekben az MI-lehetőségek[168] a hagyományos adatbázis képességekkel egyesítve hatékonyabb kereső szolgáltatást nyújtani. Az ilyen fejlődés azért előnyös, mivel organikus, hiszen felhasználja a korábbi, strukturált (relációs) adatbázis-támogatással feltöltött anyagokat.

Először a kétféle működés közötti különbséget nézzük meg egy webshop termékajánló funkcióján keresztül. Ehhez a funkcióhoz, az erre programozott ágens az emberek digitális életében generálódó adatokat veszi figyelembe a hagyományos webshop portál. Vagyis egy vásárló termékkattintásaihoz tartozó termékcímkéket a rendszer egy adatbázisban tárolja. Ha kattint az illető egy kategóriára, akkor nála az adott címkéhez tartozó pontérték megnövekszik, ha sokat kattint egy címkére (mert válogat), akkor jócskán megnő ez a súlyérték. Ajánláskor ez alapján tud a rendszer a többi hasonló (egyező címkéjű) termék közül ajánlani neki. 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. Holott valójában csak egy adatbázis aktuális állapotának lekérdezését végezte el. Az ilyen adatbázis „relációs", tehát azt a relációt, vagyis azt a valós kapcsolatot tárolja a rendszer, ami egy adott felhasználó és egy-egy terméktípus között van. Egy ilyen portál esetében az adatbázisba belépve az üzemeltető lekérdezheti ezeket a statisztikai értékeket, és értelmezni tudja azokat: pl. hogy Gipsz Jakab urat a glettanyagok érdeklik. Egy neuronháló feketedobozában (ld. II.2.1.) ezzel szemben még a legmagasabb jogosultságú adminisztrátorok számára sem értelmezhetőek a kapcsolatok és adatok, senki nem tudhatja, hogy Gipsz úr mire és mennyit kattintott.

Térjünk rá arra, hogy hogyan is néz ki ideális esetben a kétféle rendszer integrálódása. Ezt a folyamatot egy szakirodalom-kereső példáján mutatom be.

- Korábban egy írásmű feltöltésekor számos címkét (metaadatot) kézileg megadtak (szerző, témakör, évszám stb.), melyek alapján gyorsan és célirányosan lehetett információt keresni.

[168] Pontosabban itt a természetes nyelvfeldolgozásról (NLP) van szó (ld. II.3.4.).

- Ezután ennek a munkának egy részét erre programozott botokra[169] bízták, amik automatikusan végezték a címkézést. Ám ehhez a szöveget megfelelőképpen elő kellett készíteni, ezért ez sok helyen nem vált be, emellett – még ha jól is működött, akkor is – sokáig kellett (pl. a fejlődés miatt) további metainformációkat megadni.
- Később a botokat intelligens ágensek váltották le (I.2.4.), melyek egyre jobban eltalálták a szükséges címkéket.

A technológia mai állása szerint az új tartalmak címkéit egy erre betanított MI-ágens elvileg már a cikk formázása alapján is felismerhetné, és ez alapján töltené ki a szerző, a kiadó neve stb. mezőket, sőt jó esetben akár meg tudná különböztetni a kiadás évét a mű címében szereplő évszámtól is. Lassan tehát elmaradhatna a humánerővel végzett adatlapkitöltés. Ezzel szemben az általam tesztelt kutatástámogató szolgáltatások meglehetősen hiányosan végzik el ezt a feladatot. Ha működik is ilyen technológia, a már rögzített publikációkon a legtöbb online portál nem futtatja, így sok cikk máig csak kézileg tehető át pl. egy Zotero adatbázisba.

Megjegyzendő, hogy az ilyen MI-vel támogatott rendszerekben működő hagyományos adatbázisok helyett a mai kutatások egyre inkább egy neurális adatbázis irányába (II.3.5.)[204] próbálnak meg fejleszteni. Bár még korántsem 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 hasonlóan ahhoz, ahogyan a mai processzorokba integráltak számtalan, korábban különálló funkciót. Elvileg bármit részévé lehet tenni a gépi tanulásnak, ám visszafogó tényezőt jelent e téren – a biztonsági problémákon túl, – hogy megéri-e finanszírozni minden fejlesztést, nem olcsóbb-e sok téren a humánerőforrás?

V.1.3. Okosdolgok: a terminológiai anomáliák piaca

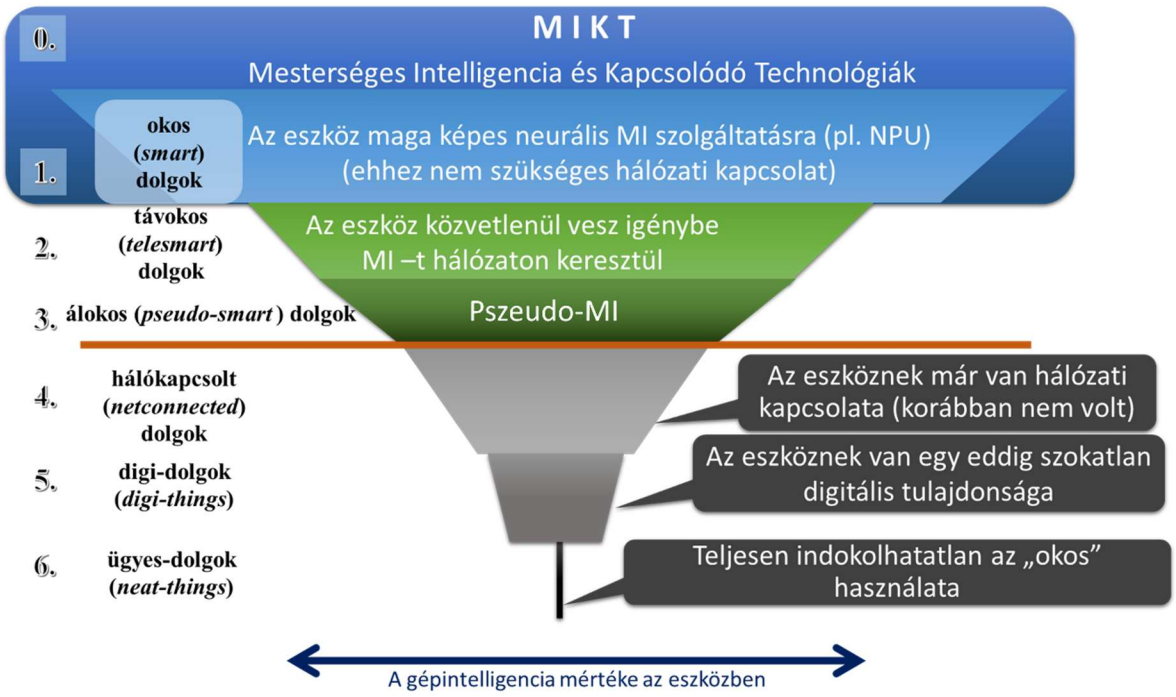
Régebben mindent a „modern” jelzővel próbáltak eladni, ma pedig bármire ráerőltetik az „intelligens” vagy az „okos” (*smart*) kifejezést. Ezáltal kezd kiüresedni az okos szó (is). Megszokjuk, hogy a boltban szinte semmit sem jelent (vajon majd az emberekre hogyan alkalmazzuk?). Bár itt az MI-fogalmat összezavaró harmadik tényező a vizsgálat tárgya, ám mivel az intelligencia kifejezésnél jobban dokumentálható az „okos” szó ilyetén használata, ezért ezen keresztül mutatom be a kérdést: rendszerezem az okos kifejezés alkalmazását (annak technikai részét), valamint javaslatot teszek pontosabb terminológiára. A 14. sz. ábrán bemutatott felosztásom a szintek tömör leírását tartalmazza, de alább ezt pontosítom és példákkal egészítem ki.

[169] Saját megfogalmazásban: a *bot* egy virtuális térben működő roBOT, a hálózaton „mozogva” végez automatizált feladatokat. Részletesebben: [203].

A példákat termékoldalakról vettem, a reklám elkerülése miatt ezeket nem konkretizálom (bárki rákeresve többet is talál). Már bemutattam az MIKT (II.1.) kifejezést, fontossága miatt pedig a következő részbe került a pszeudo-MI (V.1.4.).

Megközelítésem szerint a tanulási képesség tehet „okossá” egy gépet. Az ábra az okos jelző jogos felhasználását sárga vonallal választja el annak a zavaró alkalmazásaitól. A két alkalmazáskörön belül is megkülönböztet fokozatokat. Itt is szükséges elválasztani a neurális és nem neurális tanulást, mert most egyszerű eszközökben talán még hasonlóak a kétféle megoldás kihívásai, de hosszú távon a neurális rendszerek problémái eltávolodnak a hagyományosokétól. Az idomok szélessége is informatív, kétféle dolgot szemléltet:

- a színes idomoknál az adott technológiakör MI-képességét jelképezi;
- a szürke idomok esetében pedig azt mutatja, hogy a jövőben mennyire válhatnak „színessé”, mennyiben lehet közük az MI-hez.



14. ábra: Az okosnak hívott eszközök felosztása (saját készítés)

0. Vitán felül áll az MIKT korábban bemutatott fogalomköre, melyben fizikai és logikai értelemben egyaránt összekapcsolódnak a neurális és egyéb technológiák. (II.1.)
1. **Okoseszközök** (*smart things*): Ezek saját, tehát offline (hálózatot nem igénylő) neurális MI-képességgel rendelkeznek. Nem csupán robotokra kell itt gondolni. Egyszerűbb tanuló funkciókat egyre több eszközbe lehet integrálni. Egy ideje a hardveres MI-támogatás sem

a szervertechnológia szintje, jobb mobiltelefonok processzoraiba is integrálnak egy neurális magot, úgynevezett NPU-t, főleg a képfeldolgozás támogatásához, de az okosautók is ide tartoznak (ld. II.3.1.). Olcsóbb eszközökön inkább szoftveres képesség segíti az ilyen funkciókat (pl. csak az arcfelismerést). Az ilyen okoseszközök „még okosabbá" válhatnak, ha saját tudásukat hálózaton elérhető MI-rendszerrel egészítik ki.

2. **Távokos dolgok** (*telesmart-things*): Akkor okos, ha hálózaton keresztül csatlakozhat egy neruális MI-hez, tehát ez a távoli szolgáltatás teszi csak okossá, – csatlakozva azonban nevezhető okosnak. Csatlakozás nélkül azonban mindössze egy hagyományos számítástechnikai eszköz vagy használhatatlan. Az előbbibe sorolható egy átlagos mai asztali számítógép, az utóbbira jó példa az okoshangszóró.
3. **Álokos dolgok** (*pseudo-AI things*): Nem neurális MI-k (V.1.1.) által vezérelt technológiák. Pl. ilyen egy okosfűtés; figyelheti a felhasználó szokásait és néhány héten át tartó manuális állítgatásból megtanulja, hogy mikor és hogyan állítsa be azt. (bővebben: V1.4.)

Valójában itt illene is lezárni az okoseszközök körét, de nem ez a valóság…

4. **Hálókapcsolt dolgok** (*netconnected things*) eszközök: Bizonyos eszközök okos verziójának egyszerűen csak van valami hálózati (vagy egyéb) kapcsolata, ezzel különböztetik meg a „buta" verziótól. Például okosfésű vagy okosfogkefe mobiltelefonos applikációkon keresztül csatlakoztatható egy mobiltelefonhoz. Ez utóbbinál például pontosabb lenne a bluetooth-fogkefe kifejezés, – ám belátható, hogy nem túl jól hangzik a „kékfog fogkefe" az angol nyelvterületen... Ide sorolhatóak még az okoseszközök bizonyos kiegészítői, melyek által az eszköz jobb képességeket kap. Pl. az okostoll által jó minőségben lehet digitálisan rajzolni (de csak az azt kezelni tudó, drágább eszközre).
5. **Digi-dolgok** (*digi-things*) Ezeknél már hálózati kapcsolat sincs, csupán valami digitális tulajdonságot építenek az eszközbe, ami szokatlan, például okosövbe rejtett powerbank, vagy okosserpenyő nyelébe épített kijelző, mely jelzi, hogy jó hőmérsékleten folyik-e a főzés és jelez, ha odaég.
6. **Ügyes-dolog** (*neat-thing*): Ez a kimagyarázhatatlan kategória, megtévesztő használat, visszaélés az „okos" szóval. Az ilyen tárgyaknál még digitális funkció sincs. Pl. gumivégű pálcákat is néhol okostollnak (ld. 4. szint) titulálnak, az „okosedénynek" az az okossága, hogy műanyag fogantyúja levehető, az okosdoboz egymásba pakolható, de létezik „intelligens" mosópor is. Ez rámutat, hogy legalább irányelvekben tiltani kellene az ilyen túlzó és sokszor megtévesztő szóhasználatokat.

Röviden érdemes átgondolni, mi állhat a szó bemutatott kiüresedése mögött. Felvethető okként az okos és intelligens kifejezések piedesztálra emelése racionalista társadalmunkban. Ám

ha ez lenne az ok, akkor már a XX. század közepén is népszerű lett volna, ezzel szemben akkoriban a „modern” volt a hasonló dzsóker kifejezés. A szakirodalom is többféle okot megjelöl. Az egyik az MI-használat füllentése, amikor egyes cégek, főleg start-up-ok valótlanul állítják azt, hogy MI-t használnak.[205] 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. Végül megemlítendő az ún. „MI-effektus”. Ez nem egy hype-effektus, hanem eszerint amikor az MI elér egy bizonyos szintű elterjedtséget, akkortól gyakran elkezdik már nem intelligensnek tekinteni, mivel rájönnek, hogy nagyon elmarad az emberi teljesítménytől.[189] Ezen forrás szerzőjével nem értek egyet, mivel egyrészt erre az elmúlt 25 évben nem látok példát, másrészt ha van is ilyen, akkor is csak az újabb MI-khez képest tűnnek a régiek *kevésbé* intelligensnek.

Ezeken túl véleményem szerint a jelenség fő oka, hogy az MI szó nem csupán egy technológiát fed már le, ennél mára jóval többet jelent. Az a rendkívül erős diszrupció, ami hozzá kapcsolódik, kihat az asszociációkra is. Az okos jelző hasonlóvá vált a luxus jelzőhöz, melyek által egy dolog egy kiváltságos vagy „VIP körbe” kerül, használata különlegességet sugall. Ezért mindenki szeretné, ha terméke beletartozna abba a halmazba, vagy legalább annak holdudvarába, amit az MI képvisel. A fenti elemzés és felosztás bemutatta, hogy 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. Összegzésként leszűrhető, hogy az okos szó degradációja jól demonstrálja a fogalmi zavar marketinges oldalát, ez a gondolatmenet alkalmazható sok más esetre is.

V.1.4. Pszeudo-géptanulás, pszeudo-MI, pszeudo-autonómia

Az ML felosztásairól szólva (II.2.4) javasoltam *pszeudo-tanulás* kifejezést, egyfajta nulladik tanulási szintnek, amelynél a hagyományosan programozott rendszer a felhasználó számára a mélytanulás benyomását kelti. 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 (ld. V.1.2.). Ez a jelenség azonban vindikál egy pszeudo-*intelligencia*, vagy esetünkben inkább „pszeudo-MI” kifejezést is, ahova az így tanuló rendszereket érdemes besorolni. Erre a következő meghatározást javaslom:

> *A pszeudo-MI egy olyan rendszer, melyről a felhasználók egy részének az a benyomása, hogy hátterében neurális gépi intelligencia működik, pedig valójában egy hagyományosan programozott számítógépes architektúra.*

Az előző fejezetekben számos példán mutattam be, hogy milyen programok tartoznak ide. Úgy 8-10 éve jobb lett volna ezt vagy a pszeudo-okosságot terminológiailag megkülönböztetni, ilyen fogalmak felvetése kissé megkésett, – ám még ma sem túl késő, és nem haszontalan az általam javasolt megkülönböztetéseket bevezetni.

A két rendszer különbségét már leírtam (V.1.2.), itt nézzük, mi az, ami a két rendszerben közös. Elsősorban a „változékonyság" tulajdonságra gondolok, mely a valós idejű adattárolás fontos jellemzője, akár adatbázisról, akár adattóról van szó. Térjünk vissza a webshop példájára: ott is pillanatonként változhat egy adat. Pár óra alatt pedig radikálisan is megváltozhat pl. egy címke súlyértéke, mondjuk a vásárlók között terjedő jól sikerült reklámvideó 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.

Ezáltal viszont egy ilyen hagyományos rendszer sem lesz teljesen kiszámítható (IV.3.). Ez pedig az autonómiát érinti, ezért be kellett vezetni a pszeudo-*autonómia* kifejezést is (III.4.1., „Alfa szint"). Itt csak annyit jegyeznék meg, hogy ide kell sorolni a változásra érzékeny rendszereket akár neurális, akár hagyományos alapúak. Végezetül leszögezem, hogy ez az ál-gépintelligencia nem pejoratív vagy negatív fogalom. Egyszerűen egy bevett ógörög képzőt alkalmaztam a hagyományos számítógépek csúcsteljesítményére.

V.1.5. Az MI-fogalmat homályosító tényezők kezelése

A fenti „terminológiai degradációs modell" érthetővé tette, hogy mit is értettem az MI fogalmat homályossá tevő három tényező alatt. Ezt a három dolgot nem egyformán kell kezelni, amikor egy javított MI-fogalmat szeretnénk elérni. Alább arra teszek javaslatot, hogy hogyan lenne érdemes figyelembe venni ezeket a tényezőket.

1. Nem lehet egy javaslat által törölni a szakemberek által már régóta, széles körben és a mai napig használt nem neurális MI-fogalmakat (V.1.1.). Egyrészt, mert az ilyen megközelítések egy része mára úgyis beépült a neuronhálók fejlesztésébe, ezekben az esetekben helytelen is lenne. Másrészt az MI ilyen értelmezései szerves részét képezik a tudományos terminológiának.
2. Elsősorban oktatás által kell kezelni a felhasználók által félreértett szóhasználatot, amikor nem tudják, mit is tud, de mire nem lehet képes az általuk használt szolgáltatás (V.1.2.).

Olyan egyszerű népszerűsítő kampányokkal, mint amelyek beváltak már a kiberveszélyekre való figyelmeztetések terén, jól lehetne az ilyenfajta félreértéseket is eloszlatni. Ezt támogatná, ha szakmai szervezetek irányelveire lehetne ennél hivatkozni.

3. Szabályozandó lenne a marketing kifejezésként használt túlzó szóhasználat és a technikai tartalommal bíró pozitív szavakkal való visszaélés. Az „okosság" kifejezés esetén bemutatott marketingfogások (V.1.3.) kezelése teljesen más technológiákat is érinthet: ilyen lehet pl. a kvantum szó, az érzelem szó és sok egyéb kifejezés megtévesztő használata. Ez nem csupán elősegítené az egyértelműbb kommunikációt, de a vásárlók védelmére is hasznos eszköz lenne.

Mindhárom eset kezeléséhez javaslom a szóhasználatban előtaggal említeni, hogy a hivatkozott rendszerek mit használnak. „Indeterminisztikus MI" helyet inkább pl. „neurális MI"-ként érdemes utalni az ezt használó rendszerekre, tudatosítva, hogy nem ugyanazok a fejlesztési kihívások (IV.1.) és szabályozási problémák vonatkoznak rájuk. Viszont a „pszeudo-" előtagot kellene használni akár a tanulás, az MI, vagy az autonómia szavak előtt, (V.1.4.), amikor hagyományos rendszerre utal valaki (erre felmerülhet a „determinisztikus" előtag is, ha a kontextus alapján ez világít rá jobban a megkülönböztetés okára). Hosszabb távon ettől talán letisztul a félreérthető használatok ezen típusa.

V.2. Paradigmaváltás, de csak a szavakban: multikogníciós rendszerek igazi AGI helyett

Az V.1. alfejezetben bemutatott és osztályozott *terminológiai degradációs modell* magát az MI kifejezést nem csak jelenlegi értelmében érinti, hanem annak jövőjében még valószínűbb módon jelenik majd meg. Állítom ugyanis, hogy az AGI sokak által jósolt megjelenése a szó tartalmának eróziója által várható. Korábban bemutattam (I.2.3.), hogy régóta elfogadott szintek alapján, egy Mesterséges Általános Intelligencia (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. Azt is láthattuk, hogy a jelenlegi technológiák bármilyen bonyolultak, mind csupán az ún. Mesterséges Vékonyintelligencia (ANI) halmazába sorolhatók. A kifejezés története is igazolja állításom hátterét, miszerint „egyre lejjebb kerül a léc" abban a tekintetben, hogy meddig is mehet el ez a technológia. Elegendő ehhez vázolni „csúcs-MI" nevének és definíciójának változását.

Az AGI-t korábban hívták Erős MI-nek is (Strong AI), ám ekkor a koncepció lényege a *mesterséges tudat* elérése lett volna.[206] 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 egyszer módosult, miért ne módosulhatna többször? Ezért nem zárható 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 az olyan utópiáktól, mint a „mesterséges tudat”, vagy az „igazi, emberi megértés (vagy érzések) gépi leképzése”. Ez a „valamilyen mértékben eltávolodás” egyénenként eltérő, de mégis besorolhatóak két nagy típusba az AGI megközelítései. Mielőtt rátérnék erre a két típusra, hadd vezessek be két saját kifejezést. Azt használom fel, hogy az *omni-* előtag „minden” jelentésű, a *multi-* ehhez képes inkább a „többféle” szóval fordítható. Ezekkel megfogalmazva: a megcélzott AGI *minden* emberi kognitív képességet utánozni tudna, azaz *omnikognitív rendszer* lenne. A jelenlegi tudásunkkal megvalósítható gép azonban legfeljebb egy többféle kogníciót egyszerre utánzó, vagyis *multikognitív rendszer* lehet. Amint láttuk (I.3.3.) az emberi intelligencia-típusok száma sem lezárt kérdés még – tehát nem lehetünk képesek minden emberi kogníció- és intelligencia típus utánzására sem. Most visszatérhetünk a két féle AGI-megközelítésre, amihez a következő terminológiai felosztást definiáltam (számos leírás figyelembevételével):

1. **technikai AGI**, vagyis multikogníciós rendszer. 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,[170] nem pedig adott célokra tervezett gép.
2. **emberies AGI**, azaz omnikogníciós rendszer. Gyakran értenek az AGI alatt az ember kognitív képességeit birtokló 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 is utalnak így az AGI-ra (bár inkább csak nyilatkozatokban), ezért magas olvasottságú cikkeket és könyveket lehet e témában írni.

[170] „*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*”.[10]

Erre a felosztásra nem találtam forrást még az AGI egyik 2010 óta működő szaklapját[171] átnézve sem, amely pedig számos színvonalas írást közöl.[172]

Ahogyan erre az intelligencia vizsgálatakor (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 egy olyan szolgáltatás, 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 termék tudása messze nem lesz az emberi kogníció bonyolultságához mérhető, csupán egy modulárisan összetett vékony MI-rendszeregyüttest hívnak majd így. 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 esetleg 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 marketinghullámához 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-ként reklámozott termékekben. A „emberies-AGI" típusú, humanoid omnikogníciós rendszerről főleg nem beszélhetünk, nemhogy az első AGI-nak hívott termékeknél, de még sokáig. Számtalan 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.3.2.) egyvelege, egymás mellé rakása sem azonos ugyanezen intelligenciafajták emberen belüli összességével! Ehhez járul ezen intelligenciafajták bonyolult és ismeretlen kapcsolódásainak, egymásra hatásainak gépi megvalósítása, mely egyelőre kilátástalan.

A *szavak degradációja* egy minden „vágyott technológiára" alkalmazható modell. Erre volt példa az „okos" szó degradálódása (ld. V.1.3.) is. Az ott leírtak további bizonyítékot adnak a felvetett várakozásra, hiszen ugyanaz a mentalitás fog az AGI esetében a háttérben állni, melyet ott részletezek.

[171] https://sciendo.com/journal/JAGI

[172] Még az MI fogalmát is így közelítik meg: [207].

A KUTATÁS ZÁRÁSA UTÁNI FEJLEMÉNY

A kutatás lezárása után nem sokkal részben igazolódott is az itt felvetett állítás – bár én ezt egy-két év múlva vártam. A fenti „jóslatot" sajnos nem tudtam megjelentetni az esemény előtt, de az alábbi fejlemény előtt (2024. november 30-án) leadtam egy tanulmánykötet részére.[196]

Pár hét múlva, 2024. év végén jelent meg egy hír, melyben az OpenAI már állítja magáról, hogy megvalósították az AGI-t.[208] Ez persze elvileg lehetetlen, hiszen szó sincs multi-kogníciójáról, csupán egy nyelvi-képi szolgáltatásra húzták rá a kifejezést. Tehát az előzőekben kifejtettek szerint valóban a marketing oltárán lett az AGI kifejezés feláldozva. Jóval előbb, mint az várható volt, és egy jóval egyszerűbb rendszerre alkalmazva. Tehát még azt a fenti meglátásomat is igazolta a bejelentés, hogy még a technikai AGI sem fog az elnevezés mögött teljesülni a név használatakor. Ezzel egyébként egy harmadik meglátásomat is alátámasztották (IV.4.1.), hogy nem szabadna gazdasági síkra szűkíteni a fejlődés megközelítését. Ehelyett javaslatom szerint pontosabb lenne a horizontális és vertikális fejlődés modelljét alkalmazni. Ugyanis az OpenAI úgy fogalmazta meg bejelentését, hogy abban az AGI meghatározása még a vártnál is inkább gazdasági síkon maradt: „*az AGI elérése akkor valósul meg, ha egy MI-rendszer képes 100 milliárd dollár nyereséget termelni*" [209].

Ez a bejelentés nem változtatta meg jelentősen az AGI szóhasználatát, nem vették még át a versenytársak sem, sőt maga a Microsoft cég sem erőlteti ennek használatát. Ezért az itt említett eset úgy tűnik, egyelőre előhírnöke csupán a terminológiai degradációs modell fentebb bemutatott alkalmazásának. Ez a véleményemet csak megerősíti arról, hogy az AGI használata a mai értelemben vett általánosság előtt terjedni fog a termékek leírásában (vagy elnevezésében).

V.3. AZ MI HATÁSÁRA AZ INFORMATIKA ÉS A HUMÁNTUDOMÁNYOK IS KONVERGÁLNAK

Az MI könnyen félreérthetővé válik, ha úgy közelítek meg, hogy az intelligenciát vagy az informatikát valami objektív, szilárd dolognak tekintik.[173] Korábban bemutattam már (I.), hogy az intelligencia rugalmasan közelítendő, csak részben objektivizálható fogalom. Most az informatikával kapcsolatban vetek fel hasonlókat. Az alfejezetben feltárom, hogy bár ennek jelentése ugyan nem túl sokrétű, mégis érdemes rugalmasan állni hozzá.

[173] Jól példázza ezt a következő pár sor: „*a mesterséges intelligenciát az is meghatározhatja, hogy mit csinálnak a mesterséges intelligencia kutatói. [Ez a jelentés] a mesterséges intelligenciát elsősorban az informatika olyan ágának tekinti, amely az intelligencia tulajdonságait az intelligencia szintetizálásával tanulmányozza*".[21, pp. 14] Az ilyen megfogalmazások vezetnek az V.1.-ben tárgyalt problémákhoz.

Alább egyszerre teszek lépéseket az MI fogalmi és védelmi vizsgálata (C1 és C2) felé, mivel a kétféle szóhasználaton keresztül azt is bemutatom, hogy az MI-nek köszönhetően az *információval való foglalkozás* is gyökeresen átalakulóban van. Már korábbi kutatásaim során az a benyomás alakult ki bennem, hogy a számítástechnikában szinte kezdettől fogva az MI létrehozása volt a cél, ezért olvasta Neumann János is előszeretettel a kortárs agykutatók eredményeit. Sokáig sok minden nem állt még készen (bár megvalósult helyette valami más). A technikai szint elégtelensége (II.) mellett sokáig 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. Itt előbb megvizsgálnom, hogy az informatika kifejezés tartalma hogyan változhat meg az MI miatt, ezen belül az információ, az informatika és egyéb, ezekhez kapcsolódó humántudományi fogalmak kölcsönhatását a MI-vel.

V.3.1. Az információhoz kapcsolódó tudományok és változásuk

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, Komenczi Bertalan, egy igen alapos és széles spektrumot áttekintő egyetemi jegyzetében teszi.[210] Az információelmélet (*information theory*) alatt azonban a nemzetközi és hazai szakirodalom szóhasználatában inkább egy hírközlési fogalmat értenek. Ennek kiindulópontja és alapja egy matematikai modell [211], ami egy üzenet információtartalmára és információveszteségének kiszámítására használható. Így az „információelmélet" az információ objektív megközelítése. Jelen vizsgálódás azonban olyan interdiszciplináris megközelítést igényel, amelyben az MI újszerűsége is jól kezelhető. Ezt az elvárást az információtudomány[174] kifejezés tartama közelíti meg legjobban. Találkozhatunk az *informatológia* kifejezéssel is, mely azonban hol az információelmélet szinonimájaként, hol az információtudomány értelmében jelenik meg. Bár eredetileg kifejezetten egy metatudomány neveként kellett volna, hogy elterjedjen [212], és egy horvát folyóirat neveként is megjelent egy ideig,[175] mégsem annyira bevett kifejezés. Ezért most az információtudomány vázolására térek ki.

A könyvtártudományi irányból közelítette az információ problémáit Harold Borko (amerikai pszichológus és információtudományi szakember), és mint a pszichológia doktora kutatta az

[174] Az információelmélet markánsan megkülönböztetendő az információtudomány elméletétől.

[175] 1969–2022 között adták ki. https://hrcak.srce.hr/en/informatologia (megtekintve. 2024.07.17.)

automatikus nyelvfeldolgozás és az indexelés kérdéseit. Annak érdekében fogalmazta meg, hogy mit is ért információtudomány alatt [213], hogy az Amerikai Dokumentációs Társaságot átnevezhessék Információtudományi Társasággá.[176] Vagyis, hogy a történelmi kihívásoknak jobban megfelelő, általánosabb módon közelítsen a tudományos kutatás az információhoz. Cikkében az információtudomány[177] rugalmasan kapcsolja össze a régóta fejlődő dokumentációs tudományágat[178] 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 kezeli az információtudományt, mint a „könyvtár-büfé szakot", ám ezt a felfogást véleményem szerint az MI erősödése szét fogja morzsolni. Nem lesz elegendő ugyanis az adatok kezelésének technológiájánál leragadni, mivel az MI-ben igazi értelmet nyer az elektronikus[179] információfeldolgozás kifejezés, amely alatt eddig az 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.[180] A tudománynak máig nem született szabványos, mindenki által elfogadott definíciója. Rengeteg kérdéskört felölel, így nagyon sok alterületre osztható mind a három nagy területe [218], vagyis az elméleti, az empirikus és az alkalmazott in-

[176] A kifejezés történelmi helyzetből következő szükségességét elemzi: [214].

[177] „*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 tudományszolgáltatásokat és termékeket állít elő*" ld.: [215, pp. 314]

[178] Ennek alapjait Paul Otlet fektette le 1903-ban, aki már 1892-es esszéjében tulajdonképpen egy világ-információs adatbázis víziójával kezdett el foglalkozni. Ld.: [216].

[179] Pontosabb lehet gépi információfeldolgozásról beszélni.

[180] [217, pp. 6] (magyarul https://tmt.omikk.bme.hu/tmt/article/view/3196/4214)

formációtudomány. Katonai, illetve védelmi szempontból kiemelendő, hogy az információtudomány részei az információ védelmével, illetve biztonságos kezelésével kapcsolatos kutatások és módszertanok is. Napjainkra pedig az iménti tudományfelsoroláshoz hozzávehetjük a biológiát, az orvostudományt, a jogot, a szociológiát… Bizonyos szempontból tehát 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 információmegközelítést-, 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,[181] 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 (ld. még V.3.2.). Ő egy MI-vel kapcsolatos tudományos vita során [220] 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 irányaiként: 1. a matematikát (beleértve a formális logikát, a programelméletet és a programozási 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 az agy különböző szerveinek részletes tanulmányozását).[221, pp. 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 valamiféle hitnek.[182] Hiszen magának a megismerésnek (ezzel együtt a megismerőnek) a szubjektív oldalát kellene teljesen megszüntetni ahhoz, hogy teljesen objektivizáljuk, tudományosítsuk a kutatás tárgyát (a megismerést).

[181] A rendezvényen a tudat és tudatosság lehetséges megközelítéseit kísérleti pszichológusok, számítógép-kutatók és nyelvészek boncolgatták. Ld. [219].

[182] A kognitív tudomány lehetőségeibe vetett bizakodás hit jellegére mutat rá pl. [222].

Röviden és sarkítva: így a szubjektivitás objektivizálása az elvárás. Vagyis az ismeretelmélet régi dilemmái igazából nem oldódtak fel ezen tudomány által az elvárt mértékben.

Ide kapcsolódik az a másik probléma, ami a kognitív tudomány multidiszciplináris jellegét érinti, hogy *hogyan* képes realizálni résztudományainak egyesítését egy ilyen tudományhalmaz? Egy gyűjtőfogalomként való egyesítés nem tűnik reálisnak vagy működőképesnek, mivel ez a közös halmaz rettentően sok dolog átgondolt ismeretét igényelné. Egy multidiszciplináris tudományhalmazban csak sok oldalról magasan képzett, szintetikus gondolkodással megáldott emberek lennének képesek tevékenykedni, egyben látva át a sok résztudomány minden releváns eredményét. Tehát zseniális reneszánsztípusú gondolkodók sorát várná el egy olyan korban, ahol már a résztudományok résztudománya is túl tág annak magas szintű műveléséhez. Felmerül, hogy egy MI-rendszer válhatna-e ilyen multitudóssá. Ebben sokan hisznek, anélkül, hogy bizonyítékot láthatnánk arról, hogy a rengeteg tudomány által összehordott rengeteg tudást képes lenne egy ilyen gép valaha is *helyes* szintézissé formálni. El lehet gondolkodni, mennyire határtalan a korlátolt ember által készített gépi agy (akár a korlátolt gép által készített gépi agy). Számos további speciális kérdés is felvetődik,[183] melyeket itt nem elemzek, csupán egy megoldási javaslatot ismertetek 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 metszethalmazaként is.[184] Ez a vizsgálható kogníciós jelekre és az ezekből alkotható modellekre összpontosít. Így az érintett tudományok nem rész-, hanem segédtudományai a megismeréstannak, mely csupán használható modelleket kíván kialakítani. Az akadémikus szerint mára mindez tovább alakul egyfajta interpretált kognitív tudománnyá, mely még realisztikusabb, és a területek közötti párbeszédre alapul, konkrét problémák dialogikus megoldására törekszik.[185] Ez fontos tanulságot rejt vizsgálódásom szempontjából is, mivel rámutat, hogy ez a dialogikus megközelítés lehet a realitás, amikor az MI megismerést érintő problematikáihoz keresünk megfelelő perspektívát.

Összefoglalva: ez a terület 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

[183] Jó néhány ilyen érdekes kérdést feszeget pl. [219].

[184] Remek ábrákat is közöl a szerző az ismertetett felfogásmódokhoz. [223, pp. 1125–1126].

[185] Ez utóbbi rátekintés a kognitív tudományt interdiszciplináris oldalról ragadja meg. A multidiszciplináris oldala akkor tűnik elő, amikor egy kérdés megválaszolásához a sok területet egyszerre kívánják figyelembe venni.

világ hatékony és biztonságos virtualizációs módjainak megragadásával visz közelebb egyfaja „mesterséges megismerés” hogyan-jához. Az információtudomány „a megismerés éteri anyagát”, az információt ragadja meg, annak specifikus sajátosságaiban. Mindkettő terület fontosabb szerepet fog játszani a jövőben (az MI miatt), mint a múltban.

V.3.2. Az informatika szó tágabb és szűkebb használata

Ebben a részben feltárom, hogy egyfajta érdekes pulzálást mutat az informatika fogalom értelmezése: hol egy szűkebb adatfeldolgozást, hol pedig az információ kezelésének komplex feladatát értik alatta. Röviden: az informatika kifejezés első felbukkanásaikor még technikai értelemben használták a szót, később a tudományos világ egy része – elég korán – elkezdte jóval tágasabban is értelmezni. Aztán mégis a „számítástechnika” szinonimájaként terjedt el az informatika, ám a közeljövőben az MI talán rákényszerít minket, hogy ezt a kifejezést újra egy tágabb értelemben használjuk. Az informatikatörténet általában a *számítás* és a *technika* történeteként kerül megfogalmazásra, tehát eleve csak a szűkebb jelentést vizsgálja, ezt a pulzálást nem. Bár találtam ez alól üdítő kivételt, ami a szokásos számítástörténetre csak az információ ősi tárolása, továbbítása és a régi adatfeldolgozás ismertetése után tér rá, így egy picit tágabb értelmezés szerint közelíti, ám ez elég régi mű, [186] ezért hiánypótoló lehet, ha itt az informatika szó értelmezésének történetét tárom fel.

Karl Steinbuch német tudós, – akihez a szó első használatát kapcsolják – az „*informatik*” kifejezést az információ szó elejéből és az automatika szó végéből alkotta.[187] Erről szóló művének címe meg is adja az általa használt definíciót, vagyis hogy az „Informatika: automatikus információfeldolgozás”.[226] Itt megjegyzendő érdekesség, hogy Steinbuch behatóan 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 szórakoztatóelektronika és az adatfeldolgozás egyesülését, vagyis a multimédia korszakát.[227]

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

[186] Bár ez a kiadvány sajnos több, mint 30 éves, de még így is feleslegessé teszi, hogy külön fejezetben foglalkozzak az informatika korai történetével.[224]

[187] A pontosság kedvéért: egy Helmut Gröttrup nevű kollégájával közösen alkották a szót.[225]

mentes „informatika” szócikk.[228] Egyébként etimológiaként ő is az információ és az automatika összekapcsolását adta meg, és minden olyan tevékenységet beleértett, amely az adatok automatikus feldolgozásával kapcsolatos,[188] 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.[230]

A fogalom azonban gyorsan változásba kezdett: a keleti blokkban inkább információs szolgáltatásként lett használatos, angol területen viszont az információtudománnyal (V.3.1.) összefüggésben is elkezdték használni.[229] Több forrás szerint ez utóbbira az első példa egy 1965-ös szimpózium[189] a Kaliforniai Egyetemen, bár erről szóló jegyzőkönyvet nem tudtam fellelni. Az OECD[190] is már tágabb értelemben propagálta a kifejezést 1971-ben az „OECD Informatikai Tanulmányok” című kiadványsorozatában, amelyben úgy szerepel az informatika, mint „*az információtartalom, a reprezentáció, a technológia, valamint a felhasználásukhoz kapcsolódó módszerek és stratégiák tanulmányozása*”.[232, pp. 7]

A kortárs meghatározások közül néhány: a Collins-szótár szerint a brit angolban egyszerűen az információtudomány szinonimája, az amerikaiban az információfeldolgozás tanulmányozása és számítástechnika értelemben használatos.[233] A Cambridge-szótár szerint azonban a*z informatika* magában foglalja az információtudományt, az információfeldolgozás gyakorlatát és az információs rendszerek tervezését is.[191] Érdekes módon a magyar online értelmező szótárban nincs ilyen szócikk (csak az információra van meghatározás), viszont a magyar katonai szóhasználatban is megjelenik ez a tágabb értelmezés.[192] Felesleges a szótörténettel ennél részletesebben foglalkozni, ennyiből is érzékelhető, hogy a tágabb és szűkebb értelmezés is régóta fut egymás mellett.

Végül megemlítem, hogy a nagy múltú Edinburghi Informatikai Iskola jelenlegi saját definíciója a tág értelmezést az MI szélesíti még tovább. Ez a megközelítés nyilván összefüggésben

---

[188] 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.[229]

[189] Van, aki tévesen az informatika első felbukkanásaként utal erre: [231].

[190] Organisation for Economic Co-operation and Development, magyar nevén Gazdasági Együttműködési és Fejlesztési Szervezet.

[191] Bár nem ezzel definiálja, első meghatározásként „a*z informatika* az információkat tároló, feldolgozó és kommunikáló természetes és mesterséges rendszerek szerkezetét, viselkedését és kölcsönhatásait vizsgálja.”[234]

[192] „*Az informatika a tudomány és a technika azon területe, amely az információk keletkezésének, kezelésének és felhasználásának elméletével, gyakorlati megvalósításával és eszközrendszerével foglalkozik.*”[235, pp. 333–340]

van olyan korai, úttörő lépésekkel, hogy 1966-ban itt hozott létre egy Gépi Intelligencia és Percepció Tanszéket Donald Michie [236], valamint, hogy itt alkotta meg 1973-ban a kognitív tudomány kifejezést Christopher Longuet-Higgins.[220] Az általuk manapság használt definícióban ezért nem meglepő, hogy az informatikát egy olyan széles és változatos tudományágként definiálják, amely felöleli a számítástechnika mellett az MI-t, összefonódva a kognitív tudománnyal.[237] Az informatika tudományának ebben a megközelítésben egy kétirányú kihívásra szükséges választ adnia: (1) Feladata egyrészt annak meghatározása, hogy a természetes rendszerekből származó elvek milyen mértékben alkalmazhatók újfajta mérnöki rendszerek kifejlesztésére. (2) 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.[193] A kifejezésnek ezt a tág megközelítését kívánom én is követni, hozzátéve, hogy az informatikának azt is vizsgálnia kell, hogy a természetes rendszerekből milyen veszélyek irányulnak az információs rendszerekre, ezek hogyan kerülhetőek el (informatikai biztonság), sőt azt is, hogy milyen módon veszélyeztetik az emberi normákat (pl. az MI elfogultsága, vagy az autonómia problémák miatt, ld. IV.2. éa III.), és ezek hogyan kerülhetőek el.

V.3.3. Bővített informatika vagy új fogalom?

Az előző rész gondolatmenetét folytatva itt megnézem, hogy milyen módon lehet terminológiailag lefedni az említett tágabb jelentést. Az egyik lehetőség az informatika szó használata az imént említett tágabb értelemben, a másik lehetőség, sőt talán gyakorlati síkon valószínűbb megoldás, egy új fogalom alkalmazása a tágabb értelmezéshez, ami pl. „leképezéstan", „virtualitika", esetleg az „informatonómia" (az információban szabályosságot kereső tudomány) vagy hasonló új (jobb) kifejezés lehetne. Elsőre talán meglepő, hogy az előző rész végén az informatika szó tágabb használatát pártolom, hiszen az MI fogalmával kapcsolatban épp egy szűkebb, egzaktabb értelmezés mellett érveltem – ezért ezt kifejtem.

Annyiféle fogalom kering, amelyben az „*inform*" szógyök található, hogy még egy hasonló nem hozna tisztulást. Ezért is javasoltam más szógyökökkel új fogalmakat, amelyet viszont sokat kellene magyarázni, és szinte lehetetlen lenne elterjeszteni. Az új kifejezés ráadásul va-

[193]A 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.[237]

lójában épp arra hivatott utalni, ami az informatika eredeti (kezdeti) jelentése, ami az *automatikus információfeldolgozás*. Mert eddig a gép az embertől kapott információból leképezett *adatokat* dolgozta fel. Az MI előtt nem létezett gépi információfeldolgozás, csak automatikus *adatfeldolgozás*. Az MI azonban emberi képességekre tör, és eljutottunk oda, hogy már az információ (tehát nem csupán az adat) oda-vissza konverzióját is a gép végzi. Az ilyen gép már nem csupán a számára feldolgozhatóvá konvertált adatokat képes kezelni. Vagyis nem csupán adatfeldolgozó egység lesz, hanem most valósul meg a *valódi gépi információfeldolgozás*. Ezért, vagyis „a szó szoros értelme" miatt változhat a „*számítás*-technika" valódi „*informatikává*" – erre erőltetett lenne más szót találni. Másképpen fogalmazva: csak az ilyen szélesre tágított informatikaértelmezés képes megfelelően kapcsolódni az MI-terminushoz.

A fentebb felvetett lehetőségek közül az egyik verzió megvalósulása szinte bizonyos. De bármelyik is teljesül, az informatika (vagy a leképezéstan vagy más kifejezés) résztudományává válik az információtechnológia (IT). Az informatika továbbra is erősen technikai marad ebben a tágabb értelmezésben is, a különbség a jelenlegihez képest, hogy a számítástechnika jelentés kibővül a kapcsolódó tudományok metszetével. Tehát a Pléh Csaba által felvetett, metszetként értelmezett kognitív tudományhoz hasonlóan az informatikába is szervesen belekapcsolódik sok tudományág. Az MI kapcsán sokkal inkább interdiszciplináris jelleget ölt az informatika, mely elsősorban információtechnológia, az információtudomány és a kognitív tudomány metszetévé alakul (ezek a fogalmak részhalmazaik által kapcsolódnak hozzá). Annak érdekében, hogy ettől a túlzottan halmazközpontú szemlélettől elszakadjunk, a kutatás eddigi gondolatai alapján fogalmaztam meg egy rövid meghatározást, mely az informatika tágabb jelentését sajátos értelmezésben ragadja meg. Hasonló, a leképezés használatára utaló megközelítést lenne érdemes beépíteni egy új MI-definícióba is.

> *Az informatika a valóság elektronikus leképezéséhez keresi az egyre hatékonyabb biztonságos megoldásokat.*[194]

V.3.4. Minden tudomány konvergenciája és ennek következményei

Bemutattam, hogy a jövő tágabb perspektívát vár el az informatikától az MI miatt, ezért az interdiszciplináris tudománnyá válhat. Itt a többi tudomány irányából közelítek, hogy rámutat-

[194] Megjegyzés: A *hatékonyság* használata a fejlesztés motivációinak óriási és sokrétű halmazát igyekszik egy szóba tömöríteni. A *biztonság* szó kétirányú: a valóságos rendszerek és a virtuális rendszerek között oda-vissza érvényes.

hassak ezek közeledésére: a jelenség ugyanis jól hasonlítható *digitális konvergencia* effektusához[195], annak további folytatásaként, kiterjesztéseként is leírható. A digitális konvergencia alatt azt szokás érteni, hogy a kommunikációs eszközpaletta egybeolvad a számítógépes fejlesztésekkel. Tehát ahogyan a XXI. század elején az informatika és a telekommunikáció (telefon, TV, rádió) infokommunikációvá olvadt egybe pl. az okostelefonokban, úgy a XXI. század közepének egyik jellegzetességévé válik valószínűleg a tudományok hasonló összeolvadása, [196] elsősorban az MI a technológiával összefüggésben.

Végül rátérek arra, hogy milyen gyakorlati következményei lehetnek ennek az új informatikaértelmezésnek. Ha minden tudomány az MI-be konvergál (abban ér össze), az létrehoz egy olyan igényt, hogy egy MI-fejlesztő vagy tanítástervező szakembernek minél több kapcsolódó tudományterületen meglegyenek az alapvető ismeretei. Hiszen ütőképesebb az olyan fejlesztőcsapat, ahol minden fejlesztő tisztában van a szükséges humán, céltudományi és reál kompetenciákkal egyaránt, csak valamelyikben jobban elmélyül szakiránya szerint.

Ennek azonban pszichológiai oldalról kicsi a realitása. Az információs tudományok kapcsán említettem, hogy a reál szakemberek régen is sokkal kevésbé voltak nyitottak és elismerőek a többi tudomány iránt, mint a humán szakemberek a technika irányában. Mára ez tovább erősödött, és a köznyelv „kockának" kezdte el hívni azokat, akik számára a számítógép adta objektív közeg helyettesíti az igazi szociális kapcsolatokat. Az ilyen ember az elvont, humán jellegű (pszichológiai, filozófiai, etikai, nyelvi, társadalmi stb.) problémák iránt kevésbé fogékony. Talán azért, mert a „humán világban" nem érzi otthonosan magát, a felhozott területeken nem szerez gyakorlati ismeretet, legfeljebb a virtuális térben élve, az ottani tapasztalatai alapján alkot ilyen kérdésekben esetleg véleményt. Hiába tanítanak neki az egyetemen ilyesmit: csak a vizsgára készül fel belőle, de mélyebben sokszor nem érdekli.

Mivel azonban az informatika tágabb értelemben vett ismeretére lesz szükség a fejlesztéseknél, ezért az várható, hogy az egyéb területeken jól képzett szakemberek vértezik fel magukat szaktudással. Ez a folyamat kibontakozóban van és erősödése várható. Az ilyen irányú mozgás a 90-es évek számítástechnikai forradalmakor nehezebben valósult meg, annyira idegen volt a számítógép világa a legtöbb felnőtt ember számára. Most azonban az ötven éves generáció is már a számítógépes világba nőtt bele, a fiatalabb nemzedékek még inkább erre szocializálódtak, így a 30 évvel ezelőttihez képest nagyobb számban érdeklődnek az MI és az adatelemzés szakmai kihívásai iránt. (Persze a korosztály nagyobb része nehezen nyit ld. IV.5.).

[195] A jelenséget már 1998-tól elkezdték tudományosan azonosítani, de jelentése máig tágul.[238]

[196] Kritikusan beszél a tudományok összeolvadásával kapcsolatban a neves akadémikus, Pléh Csaba.[223]

Egy cél-MI számára csak néhány informatikus szakemberre lesz szükség, akiknek a team ad feladatot. Pl. egy orvosi MI létrehozásához sokkal több egészségügyhöz értő kolléga szükséges, mint informatikus – ilyenkor az egészségügyi szakemberek képzik vagy képeztetik ki magukat MI-ből, nem a programozók egészségügyből. Persze mindig voltak és lesznek olyan, a „reál oldalról" jött szakemberek, akik képesek elmélyedni más tudományokban is, erre számos példát hoztam az eddigi történeti részekben is. De az ilyen egyéniségek aránya igen kicsi az IT szakmán belül. Mindebből az következik, hogy elvileg érdemes ugyan az informatikus szakokon az MI-vel összefüggésben a „kapcsolódó tudományokat" is oktatni, de még inkább szükséges a más tudományokra felkészítő képzésekbe beépíteni azt a kompetenciát, mely által képesek lehetnek az MI-projektekben való részvételre, valamint a már praktizáló szakemberek számára ilyen jellegű továbbképzésekben gondolkodni – ez védelmi stratégiai szempontból hasznos megállapítás (Dicséretes, hogy az EU-ban, így hazánkban is már léteznek ilyen ingyenes képzések az AI EDIH program a European Digital Innovation Hubs hálózat jóvoltából).

V.4. JAVASLAT AZ MI MEGHATÁROZÁSÁRA

Az új MI meghatározási javaslatok előtt tekintsük át az eddigi kutatásokat.

V.4.1. Besorolási mátrix a túl sokféle MI-felosztáshoz

Egy összetett fogalom megértéséhez mindig nagy segítséget adnak annak különféle irányból indított felosztásai. Ezek inkább kiegészíteni szokták az összetett kifejezés definícióját, csak nagyon fontos eseteik válnak annak részévé. Az MI egy igen összetett fogalom, ezért így érdemes kezelni. A definiálás előtt, azzal összefüggésben át kell tekintenünk a legfontosabb felosztásait lexikonszerűen. Ennek kapcsán átgondolható, van-e olyan fontosságú, melyet a meghatározásban is érdemes lesz megjeleníteni. Azért is került ide ez az összegzés, mivel hasonló, tematikus összegzésre nem találtam minőségi forrást, pedig a szakirodalomban számtalan felosztás található.[197] Célszerűbb volt hát a szakirodalom és eddigi kutatásaim alapján saját áttekintést adni a témáról.

A felosztások még további okok miatt is praktikusak. Egyrészt az oktatás során, de azzal, hogy általuk egy konkrét MI-rendszert az eddigieknél sokkal többféle szempont szerint kategóriákba lehet sorolni, javítják a technológia szabályozhatóságának pontosságát is. Ehhez természetesen majd bele kell dolgozni az érintett felosztásokat a különféle biztonsági, jogi vagy gazdasági szabályozókba is. A besorolás segítéséhez egy látványosabb vizuális összesítést is majd

[197] Sokszor közérthetőnek szánt cikkek egymástól (talán főleg ettől [239]) vesznek át felosztásokat.

javaslok. Helytakarékosság miatt a rövid szavakból álló listák tagjait nem egymás alá, hanem egymás mellé írom (ezek máshol vertikális felsorolásokká tehetők).

(1.) A tanulás / autonómia / kogníció szerinti felosztások

Az alábbi felosztásokat korábban már ismertettem, itt csak rendszerezettebben tekinthető át.

I. Neurális háló használata alapján: neurális MI és nem neutális MI. Ezt fontos lenne figyelembe venni az MI-definícióban is.

II. A gépi tanulás architektúrája szerinti (hány neuronréteget használ a be- és kimenet között):
   1. pszeudo-tanulás (tanulás benyomását keltő hagyományos kód, nem neuronháló);
   2. gépi tanulás (ML);
   3. mélytanulás (DL).

III. Tanulási mód szerint (főbb) szintek: (1) Perceptron; (2) ANN; (3) CNN; (4) LSTN; (5) GRU; (6) RNN; (7) CRNN; (8) GAN; (9) GPT stb.

Megjegyzendő, hogy ez a lista folyamatosan bővítendő, hiszen folyamatosan állnak rendszerbe újabb és újabb megoldások.

IV. 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.

V. Hagyományos gépi autonómia 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 meghatározású gépi autonómia szintek:
   1. Nulladik szint az egyszerű, programozott automatika.
   2. Alfa szint: a pszeudo-autonómia szintje bizonyos adatbázis hátterű programoknál.
   3. Béta szinten jelenik meg a gépi tanulás.
   4. A gamma szinten jelenik meg a gépi kreativitás és empátia.
   5. Delta szint: magára hagyható komplex autonómia.
   6. Epszilon szint: emberrel szinte egyenlő gépi szabadság, illetve autonómia.
   7. Zéta szint: autonómia 2.0.

VII. A kogníciótípusok szerinti felosztás:

- Megjegyzés: Az intelligencia különböző típusait szimuláló rendszerekhez hozzávehetjük az emberiesség (humán jelleg) minden olyan szegmensét, amelyet valami módon gépileg másolni, szimulálni tudunk. Ám a fejlettség mai fokán célszerű bizonyos, jelenleg kevésbé frekventált 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, a résztvevő affektus-alaptípusok (alapérzelmek) aránya szerinti skálán.
3. **Érzékelés**. Az emberi kognícióban részt vevő intelligenciaterületek, melyeket az MI értékel ki. A természetes érzékszervek által nem fogható jeleken túl minden érzéktípusnál a gépi szenzorok tartománya sokkal szélesebb az emberinél. Az ilyen adattá alakított érzékelésekben általában mintakeresés terén használható az MI.
   a) Látás: képértés, arc- és ujjlenyomatfelismerés, képalkotás.
   b) Hallás: hang-szöveg konverter, hangaanalízis, zeneelemzés stb.
   c) Testtudat. Vegyi analízis (szaglás és ízlelés). Tapintás.
   d) Csak gépileg fogható (pl. radioaktív, elektromágneses stb.) jelek feldolgozása.
   e) Szövegben lévő hangulatok, metakommunikáció vizuális érzékelése.
   f) Agyhullámok érzékelése.
4. **Képességek**. A bemutatott képességjellegű intelligenciaterületek alapján:
   a) beszéd: beszédkonverzió, szövegértés, válaszszöveg, beszédszintetizátor;
   b) mozgás: jármű/robot autonóm mozgatása terepen, okosváros-szabályzás, orvosi videolelet-elemzés;
   c) Egyéb: a sokféle „alternatív” intelligenciaterület szimulációjára fókuszáló fejlesztés.

VIII. A tudásszintek szerinti mai felosztás

(1) ANI (week); (2) AGI (strong) - Saját felosztásom ezen belül (i) technikai AGI; (ii) emberies (humán jellegű) AGI; (3) ASI.

(2.) Egyéb MI felosztási lehetőségek

Még néhány fontos vagy érdekes felosztás a teljesség igénye nélkül:

IX. Központi működés szerint:

1. centralizált rendszer (online) – egyetlen központi gépen vagy felhőben fut, alrendszereitől 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).

X. 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 használnak. Ez persze foglalhat el nagy tárhelyet, de használata szimpla (pl. okostermosztát);
3. Emlékező[198] célrendszerek: a mélytanulás révén emlékezetük komplexebb és pontosabb, de egycélú rendszerek (pl. képfelismerő rendszerek). Ezen belül szétválaszthatóak a rövidebb (CNN, RNN) és a hosszú távú (GRU, LSTM) memóriarendszerek;
4. Komplex gépi emlékezet: multimoduláris MI, melyben többféle neuronháló (akár vegyesen egyszerű és mély) architektúra összekapcsolása által az emlékezet kezd az emberi emlékezet sokféleségéhez hasonlóan sokrétűvé válni. Ezen belül további felosztásokat a felhasználás eredményezhet (pl. egy beszédértő hangulatfelismerő rendszer).

XI. Felhasználási mód szerint:

- Talán a legnépesebb felosztást ezek a felosztások jelentenék (pl. egészségügyi, ipari, művészeti, közigazgatási, hadi, rendészeti stb.), ezért ezt nem részletezem.[199]

XII. Átláthatóság szerinti felosztás: ennek mérése még kihívás, de említendő tényező.

XIII. Kockázat szerinti felosztás (pl. az EU hivatalos besorolási módszere [240, pp. 5.2.2.]).

[198] Több írásban a 3. szintet az Elme-elméleten alapuló rendszer (Theory of Mind AI), a 4.-et az Öntudatos rendszerek (Self-aware AI) kapják – ezeket azonban nem memóriájuk jellemzi, ezért elvetettem.

[199] Egy több ezer elemes, sok szintű listában lehetne összegezni azokat a gyűjtőterületeket, alterületeiket és azok részterületeit, amire különböző célrendszerek alkalmasak, hiszen más technológiákkal karöltve az MI mai képességei valóban az életünk minden területén használható.

- Az alábbi lista egy szabályzásorientált megközelítés, a szintek nevei 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;
3. Korlátozott kockázati kategória;
4. Minimális vagy semmi kockázatot nem jelentő MI.

XIV. Biológiai architektúrával való kapcsolat szerint:

1. csak szoftveres (hagyományos architektúrán);
2. célhardverrel is támogatott szoftver vagy hardveres cél-MI;
3. állati vagy emberi agyba implantált, közvetlenül kiterjesztve adott agyi képességeket.

Technikai nehézség a besorolások objektívvé tétele, ehhez azonban alkalmazhatóak a máshol már bevált ipari módszerek. Sőt, ezen a téren véleményem szerint reális nemzetközi szabványmegállapodásokban konszenzusra jutni, nem úgy, mint etikai vagy korlátozási kérdésekben. Természetesen egy ekkora terület esetében számtalan további felosztási lehetőség merülhet fel: tesztelhetőség szerint, elterjedtség szerint, energiafogyasztás -szerint stb., a szakirodalomban fel is merül jó néhány egyéb felosztás. Sokszor összekeverednek, egybemosódnak a különféle felosztási szempontok.[200] A besorolási mátrix szükségességének belátásához azonban úgy vélem, elegendő ennyi féle felosztás bemutatása is.

(3.) A besorolás mátrix felvetése

A besorolás úgy történhet, hogy számkódokat rendelünk a különféle besorolásokhoz. Ez a számkód megkülönbözteti a szempontokon belül az altípusokat, de az adott szolgáltatás minőségére utaló érték is lehet. Persze nem kell, sőt nem is lehet minden ismertetett besorolás szerint egyszerre értékelni egy rendszert. Bizonyos adatok úgyis függenek egymástól, tehát nem informatív minden megadása. Nézzük meg egy példán, hogy a vizsgált robot besorolását lekérdezve milyen adatokat kapnánk. Gépünk szövegértése 7-es szintű, felügyelt módon tanított (=4), tisztán szoftveres alapú (=10), centralizált (=2), minimálisan átlátható (=1). Képfelismerése eközben 8-as minőségű csavarokon, 7-es alátéteken (stb., akár alkatrész-kategóriánkként külön), felügyelet nélkül tanított (=3), hardvertámogatott (=7, ez az NPU-kódja), rajintelligencia alapú

[200] Pl. a fenti elméleti felosztás és némely felhasználási mód keveredik az elsőre látványos ábrában ebben a tanulmányban: [241, pp. 4].

(=3, fokozata 5), minimálisan átlátható (=1). Ugyanígy kezelnénk a szöveggenerálását, a mozgását és a többi képességét.

Csakhogy ha ezeket egyetlen számértékben akarjuk kifejezni (ahogyan az EU besorolása), akkor az az eredményszám nagyon csalóka képet fog mutatni a robotunkról, bármilyen matematikai trükkel súlyozva számítjuk is ki. Ha pedig csalóka, akkor igazából csak áleredmény, használhatatlan. Ezért eleve meg kell próbálni úgy megjeleníteni az eredményt, hogy az egyből felfedje a részértékeket is.

Erre kézenfekvő valamiféle vizualizáció, azonban a régebbi „exceles" diagrammtípusok kevésnek bizonyulnak ennyi paraméter figyelembevételéhez. Megpróbáltam például egy radardiagrammon csak a legfontosabb besorolásokat felvenni (nem az előbbi robotos példa, hanem egyéb kitalált adatok alapján). A 15. számú ábrán a kogníciós besorolás értékelését kíséreltem meg, de nem a szövegbeli példa alapján, hanem a látványosabb eredmény kedvéért, annál több paraméterrel. A kapott ábra az egyetlen értéknél kifejezőbb, de korántsem eléggé informatív.

A mért és figyelembe vett adatok összességét egyszerre lenne jó látni és összehasonlítani.

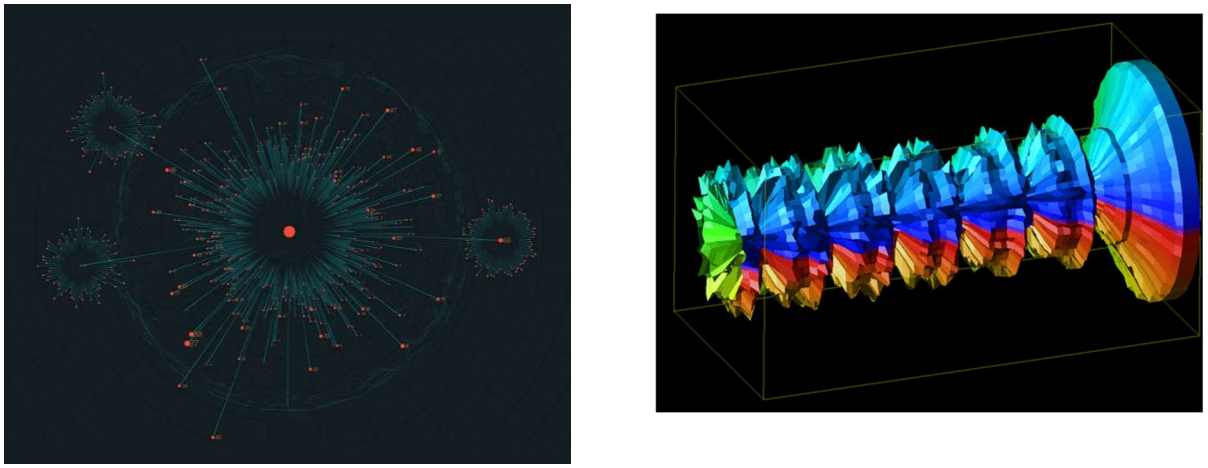
Gondolkodás	2
Érzés - output	5
Érzés - input	6
Érzékelés: látás	7
képértés	2
arc- és ujjlenyomatfelismerés	3
képalkotás	6
Érzékelés: hallás	2
hang-szöveg-konverter	8
hanganalízis	1
zeneelemzés	5
Érzékelés: szaglás	4
Érzékelés: ízlelés	3
Érzékelés: tapintás	6
Képességek: beszéd	3
beszédértés	5
szövegértés	8
válaszszöveg	2
beszéd-szintetizátor	5
Képességek: mozgás	3
jármű/robot autonóm mozgatása terepen	6
okosváros-szabályzás	7
orvosi videolelet elemzés	4
egyéb intelligenciaterület	2

15. ábra: Besorolásszintek- exceles ábrázolása (saját készítés)

Ehhez az adatelemzésben már bevált adatvizualizációs módszereket érdemes inkább használni.

Erre nem készítettem ábrát, de demonstrálásához a 16. ábrán látható kétféle módszer is, melyeket adatelemző oldalakról vettem. Mindkettő egy színezett, háromdimenziós (forgatható) tüske diagram, az első gömbszerű, a második hengerszerű elrendezésben. Ezeken a különböző besorolások alapján kapott számértékek egy-egy tüskeként állnak különféle irányokba, a kategóriákat pedig színezések kapcsolják össze (ez inkább a hengerszerűn látható). Az első ábra

nem igazán jó, ideálisabb lenne, ha az besorolási értékekből „egy színes 3D-s tüskegömb” (gömbszerűség) jönne létre, mely forgatva megmutatná az adott rendszer paraméterhalmazát. Mivel ehhez szemléltetést nem találtam, egy kétdimenziós változaton[201] mutatom be. A hengeres változathoz talált kép jobban demonstrálja a módszer használhatóságát, de mindkét fajta vizualizáció sok szempontot képes egyszerre ábrázolni. Azonban mindehhez hozzá kell tenni, hogy egy rendszer akkreditációjához nagyszámú vizsgálatra, tesztre és mérésre lenne szükség, mely nyilván drágítaná a szolgáltatást – ám nem ez jelenti az igazi problémát. Ez egy elterjedt és megszokott ártényező: ez alapján kerülhet forgalomba sok hagyományos termék is. Sokkal inkább a mérés és annak objektivitása a nagy kihívás. Mivel ezek alaposabb kifejtése jelenlegi vizsgálataimat nem segíti, nem elemzem a kérdést mélyebben.



16. ábra: Az MI besorolás-mátrix vizualizációi (szemléltetés, forrás: internet)

V.4.2. A fogalmakból hiányzó tényezők és hatásuk

Itt röviden összegzem az I-V. fejezetek alapján mely kifejezéseket vélem szükségesnek figyelembe venni annak érdekében, hogy egy teljesebb MI-definíciót lehessen alkotni. Három halmazba soroltam a tényezőket: ++ jellel jelöltem azokat, melyeket fontosnak tartok beépíteni egy új meghatározásba, + jellel azokat, amelyeket javaslok figyelembe venni, és – jellel azokat, melyek nem szükségesek, és - - jellel azokat, melyek inkább kerülendők a világos fogalmazás érdekében.

A 3. sz. táblázat választ ad K5 kérdésre, és mivel egyben a H1 hipotézis bizonyításához is jól használható, ezért minden megjelenítendő kifejezéshez megadtam az azt tárgyaló fejezeteket, illetve a részkövetkeztetés számát, amely hivatkozható lesz (a tanulmány végén) a

[201] az ábrák forrása 15. ábra: https://www.vectorstock.com/royalty-free-vector/big-data-circular-visualization-futuristic-vector-19630392, 16. ábra: https://datavizcatalogue.com/blog/further_exploration_2_3d_chart/ (Letöltve: 2024.07.01)

bizonyításban. Továbbá hasznosnak ítéltem egy oktatási oszloppal is kiegészíteni az összesítést: ez a tudományos eredményhez nem kíván hozzáadni semmit, viszont támpontot ad ahhoz, hogy ezen fejezetek mely részeit érdemes a katonai felsőoktatás részévé tenni jelen dolgozat tananyaggá alakítása során (az ide kerülő értékelés szubjektív). A + jel itt azt jelenti, hogy „megemlítendő, ha van rá óraszám, akkor magyarázandó” a ++ jelzést pedig „valamilyen szinten mindenkinek magyarázandó” értelemben használom.

	Összegzés	kifejezés	MI-def.	oktatás
1.	**Intelligenciafajták.** Több mai hivatalos meghatározásnál kimutattam (I.2.), hogy azok az ész (gondolkodás, megismerés, tanítás, autonómia kulcsszavak) körül fogalmazódnak meg. Így túlságosan is észcentrikusak, nem domborodik ki az intelligencia sok egyéb fajtája, holott azok gépi szimulációjára folynak a kutatások. Ezt az affektív számítástechnika bemutatásával (I.3.) demonstráltam. Az új meghatározásnak elsősorban ezt a hiányosságot kell javítania. (Megjegyzés: a magyar megközelítés ezt a hibát elkerülte. (I.1.2.))	intelligen-ciafajták (R1.1.)	+ +	+ +
2.	A nyelvi elemzés (I.3.1.) alapján nem került elő olyan jelentés, melyet be kellene építeni az új fogalomba. A legtöbb jelentésréteg megjelenik, a hiányzók (pl. megértés, felismerés) megjelenítése nem indokolt.	szógyök-elemzés	-	-
3.	Az **MIKT** fogalmát (II.1.) is bevezettem, mely megkülönbözteti az MI lényegét azoktól az igen összetett rendszerektől, melyek azt „körbeveszik”. Ezt érvényesíteni kell az új fogalomban, ha nem ezzel az általam javasolt betűszóval, akkor másképp.	MIKT (R2.1.)	++	++
4.	A digitális ökoszisztémát (II.1.4.) nem szükséges a fogalomba beépíteni, mivel csak következmény.	digit. öko-szisztéma	-	+
5.	A tanulási modellek és gépi tanulás felosztásai általános meghatározáshoz nem szükséges technikai többletet közölnek (II.2.),.	neurális modell-típusok	-	+

	Összegzés	kifejezés	MI-def.	oktatás
6.	A **rajintelligencia** fontosságában és egyediségében kiemelkedik a biológiai ihletettségű rendszerek közül: rámutat a decentralizált MI-rendszerek fontosságára. Pontosabb lenne egy utalás a „biológiai együttműködések digitalizálása” (II.3.3.) megfontolandó – valamelyiket bele kellene építeni).	rajintelligencia / biol. együttműködések (R2.2.)	+ +	+
7.	Az MI hardveres háttere, a neurális adatbázisok, sőt az NLP sem releváns a definícióban (II.3.1&4&5.)		-	+
8.	A jövőbe mutató labormegoldások beemelése a jelenleg célzott fogalmat zavarossá tenné, kerülendő (lista II.3.1. végén). Pl. a kiterjesztett ember megoldások a „mesterséges” szó újragondolását tennék szükségessé.	a jövőben lehetséges technológiák	-	+
9.	A **szinergia** (szinergikus vagy szimbiotikus MI) utal arra a célra, hogy olyan rendszerek jöjjenek létre, melyekkel jó együtt élni (I.2.1. & 2.3-4.), belefogalmazása javasolt – utalva arra, hogy az ilyen gép nem az ember ellen készül.	szinergia (R2.4.)	+	+
10.	**Autonómiafelosztások**(III.): Nem szükségesek, sőt ajánlott az autonómia kifejezés kerülése. Használata csak az automatika szóval együtt említve, tagadva javasolt.	autonómia	- - (+)	+
11.	**A vékony MI valójában csak automatika**: Erre utalni kell, a félelmek preventív elkerülése érdekében (III.3.2. & III.4.1.).	automatika (R3.6.)	++	+ +
12.	**MI-kihívások** (IV.1.): Egyik sem tartozik a meghatározás lényegéhez, csupán az MI leírásához.	kihívások	-	+
13.	**MI-torzítások** (IV.2.): részletezésük nem tartozik hozzá.	torzítások	-	+
14.	**Kogníció / kognitív**: a szó megjelenítése javasolt a technológia korlátaira való utalásként (III.5 és IV.2.2.)	kogníció (R4.3.)	+	+
15.	**Leképezés**: a szó megjelenítése javasolt a technológia korlátaira való utalásként (III.5 és IV.2.2.), továbbá az autonómia az informatika definíció (V.3.1.) alapján is.	leképezés (R4.3.)	+	+

	Összegzés	kifejezés	MI-def.	oktatás
16.	**A hibázás jellege** (IV.3.)**:** Bizonyos esetekben hasznos lehet a megjelenítése egy meghatározásban is, de alapvetően nem hiányzik, ha kimarad.	hibázás	-	+
17.	**Az MI történeti regressziói** (IV.4.): nem szükséges		-	+
18.	**Az MI emberi befogadhatósága** (IV.5.): Csak olyan esetben javasolt, amikor egy kurzus elején a technológia elsajátításának emberi kihívásait ki akarják emelni.		-	-
19.	**A nem neurális MI-megoldások** (V.1.) külön említése fontos, mivel annak homályosságát oldja abban a tekintetben, hogy determinisztikus vagy nem determinisztikus MI-t értünk a fogalom alatt. Valamelyik javasolt előtag megjelenítése elengedhetetlen, ha a meghatározást biztonsági szabályokhoz fogalmazzák meg.	neurális és nem neurális (in-/determi-nisztikus) MI (R5.5-A)	+ +	+
20.	**Pszeudo-MI** (V.1.4.): A definíciót nehézkessé tenné, és magyarázkodásra késztetné, ezért nem szükséges. Ám egyéb szövegekben, oktatásban, szabályzásban elengedhetetlennek tartom a fogalmak összemosódásának elkerülése érdekében.		-	+
21.	**Számítástechnika**[202]: megemlítése fontos, amikor a meghatározást biztonsági szabályokhoz fogalmazzák meg, mivel nem informatikai MI biztonsága is lényegileg tér el. (Olyan esetben lehetne tőle eltekinteni pl., ha biológiai MI-re is érvényes megközelítést akar valaki, ám ez nem a jelen vagy a közeljövő realitása.)	Számítás-technika (R5.6)	+	+
22.	**Az informatika szó kerülését** javaslom, mert amint kimutattam, az informatika szót épp az MI fényében nem pontos a számítástechnika szinonimájaként használni (V.3.1.).	informatika (R5.4)	- -	+

[202] A számítástechnika kifejezés az informatika szóval való összevetés miatt került csak bele, nem újdonsága miatt.

	Összegzés	kifejezés	MI-def.	oktatás
23.	**A tudományok konvergenciája**, mely az MI hatására létrejön (V.3.2.), bár rendkívül fontos, de fogalmilag ez sem érvényesítendő.	tudomány-konver-gencia	-	+
24.	Az **MI-felosztások** (V.4.1.) nem kell, hogy a meghatározás részét képezzék, mivel azt kiegészítik.	MI-felosztások	-	+

3. táblázat: Az új MI-definícióban megjelenítendő tényezők (saját készítés)

Az összesítésből látható, hogy számos kifejezést azonosítottam, melyeket meg kell vagy érdemes megjeleníteni egy MI-fogalomban – de ez nem jelenti, hogy az öt fejezet során felvetett számos egyéb kifejezés vagy felosztás ne jelenhetne meg benne.[203] Ezek indoklását ugyan bele tudtam tenni a fenti táblázatba, de szükséges lenne azt is bemutatni, hogy milyen szempontból érthetőek félre a megjelölt kifejezések nélkül a jelenlegi MI-fogalmak. Erre az áttekinthetőség érdekében külön összefoglalást szerkesztettem (4.sz. táblázat). Ebbe a kerülendő kifejezések nem kerültek bele, mivel legtöbbet csupán bonyolító jellegűnek ítéltem, aminek pedig erősen javasoltam kerülését, ott a fenti indoklás egyben a várható félreértést is tartalmazta.

hiányzó kifejezés	okozott félreértés
intelligencia-fajták	Egy emberértelmezési (filozófiai-antropológiai téren jelentkező) hiba az intelligencia okosságra szűkítése, mely az emberutánzás perspektíváit teszi félreérthetővé – holott az intelligencia egyre több fajtáját egyre alaposabban elemzik, hiszen ezek is fejlesztendőek minden emberben is, nem csupán az értelmi képességük. (Ez túlmutat a technológián, hiszen azt sugallja, hogy az emberi méltóság az okossággal van egyenes arányban, és a jövőben kevésbé okos polgárok megvetését generálhatja.)
MIKT	Annak félreértése, hogy egy adott rendszer egyéb technológiákkal alkot egységet, biztonsági szempontból kockázatot jelent, a használatban pedig a lehetőségek rossz felméréséhez vezet.
rajintelligencia	Egy okosentitás-halmaz felépítése és használhatósága pont olyan mértékben különbözik a központi szerveren futó MI-modelltől, mint egy

[203] Az alábbi első javaslatomban megjelenik többek között pl. az optimalizáció, a kreativitás, a szoftveres-hardveres és centralizált-decentralizált felosztások.

hiányzó kifejezés	okozott félreértés
	hangyaboly egy tigristől. Persze mindkettő „állat” – de működésüket és célelérésüket tekintve alapvetően mások, ahogyan erőforrásigényüket is eltérően szükséges kezelni.
szinergia / szinergikus MI	Ennek a jelzőnek a megjelenítése az MI-től való félelem félreértését hivatott eloszlatni (vagy legalább mérsékelni). Az embert komplementer módon kiegészítő technológia olyan cél, melyet érdemes minden érintett számára alapvető törekvéssé tenni fejlesztői, felhasználói, szabályozási, ellenőrzési vagy értékelési szempontból egyaránt.
automatika	Gyakori a gépi autonómiától való félelem, ami talán ellensúlyozható ezen másik kifejezés hangsúlyozásával.
kogníció / kognitív képességek	Az intelligencia mögött lévő világmegismerésre utaló tudományos kifejezés jól utal arra, amerre a technológia halad, hogy az emberi megismerés minél átfogóbb spektruma legyen a gép által feldolgozható.
leképezés	Fontos megjeleníteni, hogy az informatika és az MI a világot képezi le valamilyen módon gépi modellekbe, tehát a kifejezés egyszerre utal a gép és az ember elvi különbségére és a gépiesítés határaira – ezáltal mind a félelem eloszlatása, mind a biztonság szempontjából hasznos a megjelenítése.
számítástechnika	Biztonsági félreértést generálhat a jövőben. Ez a szó ugyan nem hiányzik a jelenleg élő fogalmak nagy részéből sem, viszont elhagyása felmerülhet az egyéb mesterséges kognícióbővítési technológiákkal összefüggésben – amelyekre azonban más kifejezést kellene keresni a jelentősen eltérő szabályozhatóság miatt.
nem neurális MI (determinisztikus MI)	Sokszor okozhat biztonsági félreértést, hogy nem világos a neurális feketedobozos ágensek léte vagy szerepe egy rendszerben, tehát valamilyen módon utalni kell ezek hiányára (pl. a két javaslat egyikével).

4. táblázat: Miért félreérthető egyes kifejezések nélkül az MI fogalom? (saját készítés)

V.4.3. Az új MI-meghatározások korlátai

A terminológiai vizsgálódások végén azt is fontos tudatosítani, hogy elméletileg lehetetlen egy „tökéletes” MI-definíció megalkotása. Feloldhatatlan elvi korlátokba ütközünk, többek között a következők miatt:

- **Túl gyorsan változik minden technológia.** Az MIKT terén folyamatosak az újdonságok bejelentései, melyek megállás nélkül új kihívásokat is generálhatnak. (Elég csak a ChatGPT megjelenése utána EU-s reakciókra utalnom.[242]) Ennek egyik következménye, hogy folyton változik a definíció tárgya. Az alábbi meghatározást sem terveztem hosszú időre. Ezt tovább nehezíti, hogy ha egy konszenzuson alapuló fogalmat (vagy szabályozást) szeretnénk, az könnyen okafogyottá is válhat, mire létrejöhetne.
- **Túlságosan új az emberszerűség ilyen formája.** Kiforratlansága miatt akkor is változhatna sokat a fogalom, ha nem lenne fejlődés. Jó esetben pl. a jelenlegi „nagyobb hatékonyság" paradigma helyébe emberibb mottó kerülhet, de új fogalom igénye is megjelenhet, például az MI-ellenes félelmen alapuló kampányok miatt.
- **Túl sokféle az MI**. Mint láthattuk, az utánzás számos típusa beletartozik ebbe a körbe. Ám az ember és az élővilág különböző intelligenciarészeit, vagy az érzékelési módjainkat nagyon eltérően kell a gépekbe leképezni. Ezen technológia részei (moduljai, fejlesztési irányai) között sokszor lényegesebbek az eltérések, mint a klasszikus számítástechnika eltérő moduljai között – ezt egyetlen fogalom nem mindig képes kezelni.
- **Túl sok minden tartozik és tartozhat bele az MI-be.** A homályosító tényezők bemutatása is rávilágított, hogy mivel egy 80 éves fogalomról van szó, azt nem lehet rendeletileg gyorsan megváltoztatni. Bár törekedni kell a kifejezés használatának szabatos megoldásait közismertté tenni, az emberi tényezők (szokások, tájékozatlanság, üzleti érdek) könnyen meghiúsít egy-egy ilyen nemzetközi törekvést is. (Pl. az EU meghatározza valahogy, egy techóriás pedig nem úgy használja és reklámozza – az utóbbi lesz a meghatározóbb.)

V.4.4. A javasolt MI-fogalom

Érdemes több, különböző alaposságú és célú megfogalmazást adni. Én itt kétféle irányra mutatok be példát: egy általános, összetettebb és egy oktatási célra szánt, egyszerűsített definíciót állítottam össze.

A mesterséges intelligencia a kognitív képességek egy körének leképezése egy számítástechnikai rendszerbe. Az így kapott architektúra egyes moduljai képesek az emberi intelligencia egy-egy területén a tanulás meghatározott szintjét megvalósítani, főleg az értelem és az érzelmek (illetve ezek affektusai) terén, de a biológiai együttműködések digitális utánzásával is. Elsősorban a neurális MI-rendszerekre jellemző alaptulajdonság a tanulási képesség valamilyen foka (bár ez a nem neurális MI-k

egyes típusaiban is megjelenik). A tanult adatok alapján az MI képes indeterminált, időben (a tanulás folytatása miatt) változó, rá jellemző sajátos döntésre, válaszra vagy kreativitásra is. Ezáltal a programozott, determinisztikus reakcióikon túl képes az automatika magasabb szintjeit megvalósítani, mely akár gépi autonómiának is tűnhet. Az MI-rendszerek szoftveres vagy hardveres megvalósulásai egyaránt a legnagyobb optimalitásra törekszenek, ezzel együtt (ideális esetben) komplementerként, szimbiotikusan egészítik ki az emberi képességeket minden téren, ha a felhasználó is megtanulja az ilyen rendszereket partnerként kezelni. Az MI általában robot- vagy hálózati szolgáltatásokban valósít meg ki- és bemeneti funkciókat. Inkább központi szolgáltatásként használatos, de feladattól függően egyedi egyszerű (olcsó) entitások közösségi tudásában (pl. rajintelligencia) implementálva, decentralizáltan is működhet. Robusztus központosított rendszereknél inkább MIKT-rendszerről kell beszélünk, (azaz Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák rendszeréről), ugyanis hatványozottan nagyobb tudású egy, a felhőben működő MI szorosan összefüggő rendszert alkotva IoT szenzorokkal, adattavakkal (strukturált és strukturálatlan adatokkal) és egyéb infokommunikációs technológiákkal.

A fenti definícióból egyszerűsített, bevezető kurzusokon vagy publikációkban használható MI-meghatározásom a következő:

A mesterséges intelligencia a kognitív képességek egy körének leképezése egy számítástechnikai rendszerbe, mely ezáltal képes az emberi intelligencia egy vagy több területén a tanulásra. A programozott viselkedési szinten túl, elsősorban az értelem, az affektusok és egyéb intelligenciatípusok utánzásával képesek az ilyen gépek egy magasfokú automatikára, netán bizonyos kreativitással is reagálni a világ változásaira. Akár hálózati szolgáltatásként, akár robotokba építve, az emberrel lehetőleg szimbiotikusan együttműködve törekszik a legnagyobb optimalitásra minden téren, ha a felhasználó is megtanulja partnerként kezelni. A nagy hatékonyságú rendszerek MIKT -architektúrában működnek, ahol az MI egyéb infokommunikációs technológiákkal alkot egy rendszert, de rajintelligenciakén kis teljesítményű elektronikus (akár más MI) entitások is együtt tudnak működni egymással.*

**azaz Mesterséges Intelligencia és Hozzá Kapcsolódó Technológiák*

V.5. RÉSZÖSSZEFOGLALÁS: FOGALMI VIZSGÁLATOK ÉS JAVASLATOK

Összegzés. A H1 hipotézis részben igazolást nyert az első fejezetekben, de a fenti vizsgálatok tették a bizonyítást teljessé. A vizsgálat elején a hangsúly az MI-ből hiányzó aspektusokról az ellentmondásos jelentésekre tevődött, ugyanis áttekintettem, hogy hányféle nem neurális (determinisztikus) rendszerre is használják az MI kifejezést. A klasszikus és neurális rendszerek technikai egybeolvasztása nagyon előremutató, ám súlyos félreértésekhez (akár biztonsági problémákhoz) vezethet, amikor az MI szó használatában mosódik össze, hogy neurális rendszerről van szó, vagy sem. A vizsgálatokat ennek áttekintésével kezdtem, rendszerezve azokat az eseteket, amikor az MI kifejezést nem neurális rendszerekre alkalmazzák. Három esetét különböztettem meg: a kutatók, a felhasználók és a piaci szereplők helytelen szóhasználatát vizsgáltam. Külön elemeztem a három féle probléma kezelhetőségét egy új MI fogalom kialakításakor. Ezekben a vizsgálatokban felvázoltam egy „terminológiai degradációs modellt" is. Ezt egy általánosan alkalmazható elvnek tartom, melynek segítségével akár jövőbeni pontatlan szóhasználatok is prognosztizálhatóak. Erre példaként az (egyelőre nem elterjedt) AGI kifejezésre is alkalmaztam a modellt, és így olyan eredményre jutottam, melyet időközben a valóság is részben igazolt.

Ezután lefolytattam az informatika szó tartalmának változásvizsgálatát az MI fényében. Ebből a C1 célhoz, pontosabban az új MI-fogalom számára is fontos adalék született (az informatika szó kerülése), de egyben a védelmi probléma (P2) tárgyalását is segítettem. Ugyanis számos, a témához kapcsolódó kifejezés magyarázata mellett vázoltam a tudományok konvergenciájának jelenségét (az MI hatására), ennek pedig az informatikai képzésre nézve is lettek konklúziói. A következő alfejezetben három összesítést éreztem elhagyhatatlannak: először a különböző MI-felosztások lexikonszerű felsorolását adtam közre, mely hasznos gyűjtemény lehet minden érdeklődő számára, és ezen bemutattam egy besorolási mátrix javaslatát is, mellyel sok ilyen felosztás együttes figyelembevétele alapján lehetne egy adott rendszer tudását egy palettán elhelyezni és vizualizálni.

Ezt követően summáztam az eddigi fejezetek vizsgálataiban feltárt, az MI-kifejezésből hiányzó vagy azt összezavaró megfogalmazásokat. Röviden kitértem arra is, hogy miért nem lehet tökéletes definíciót megfogalmazni egy olyan problémás jelenségnek, mint az MI. Bár nem képezte a vonatkozó hipotézis részét, és tudományos eredménynek sem tekintem, de ennyi vizsgálat után illett saját javaslataimmal növelnem az MI-definíciók már eleve jelentős számát, – bemutatva, hogy lehetséges mondatokba önteni a vázolt összegzéseket. Ezzel zártam le az H1 igazolását, és az ehhez kötődő kutatást.

P1-gyel és P2-vel egyaránt kapcsolatos következtetések

R5.1.: Az MI miatt a mai informatika kifejezés kiszélesedése várható (V.3.2-3.).

R5.2.: Minden tudomány konvergál egymás felé az MI-ben (V.3.4.).

R5.3.: A terminológiai degradációs modell szélesebb körben használható, főleg a technológiai fogalmak erodálódásának követésére (V.1-2.).

P1-gyel kapcsolatos következtetések

R5.4. A jelenlegi MI-fogalom több oldalról homályos és túlterhelt:

- R5.4-A a szakemberek régóta szívesen alkalmazzák a hagyományos számítástechnika bizonyos specifikus alkalmazásaira is az MI-fogalmat (V.1.1.);
- R5.4-B a felhasználók gyakran keverik össze a hagyományosan programozott és a mélytanuló szolgáltatásokat (V.1.2.);
- R5.4-C a piaci verseny erősíti a félreértelmezést, mivel az „intelligens” és az „okos” szavak jól csengenek akkor is, ha nincs mögöttük technikai tartalom (V.1.3.).

R5.5. Az MI-t befolyásoló fogalmi tényezők kezelése sokféle (V.1.5.):

- R5.5-A Nem lehet az MI-körből kizárni az R5.4-A szerint a szakemberek által régóta és széleskörűen használt nem neurális MI-fogalmakat, ezért a definícióban is ennek a szétválasztásnak a megkülönböztetése szükséges.
- R5.5-B Szabályzással, pontosabb új terminológiák oktatásával kezelendő a neurális és az adatalapú felhasználói szolgáltatások R5.4-B szerinti szóhasználata.
- R5.5-C Szabályozni szükséges az R5.4-C-ben kimutatott, technikai tartalommal bíró szavak piaci használatát (nem csupán az okosság és az intelligencia kifejezéseket).

R5.6. Az R5.1-gyel összefüggésben (V.3. alapján) egy jelenlegi MI-fogalomban az informatika kifejezés kerülését, és inkább a számítástechnika szó használatát javaslom.

R5.7. A kétféle MI-definíciójavaslatomat- R5.1-2-3-6-7-9. alapján fogalmaztam meg (V.4.4.), ezáltal kimutattam, hogy lehetséges az MI kifejezésnek olyan meghatározásokat adni, melyek még áttekinthető méretűek, a korábbiaknál több tényezőt vesznek figyelembe, és tudatosabban kerülnek zavaró kifejezéseket.

P2-vel kapcsolatos következtetések

R5.8. Az R5.3. (a terminológiai degradáció jelensége) miatt várhatóan az Általános MI (AGI) kifejezést jóval hamarabb használni kezdik a cégek saját vékony MI-termékeikre, minthogy valóban általános célúnak nevezhető lenne (V.2.).

R5.9. A modellek működőképessége „előreszaladt” a humántudományokhoz, illetve azok matematikai formalizálásának üteméhez képest, ezért a technológia biztonságos mederben tartásához szükséges, hogy bevárja ezeket (V.3.1.).

R5.10. A felsőoktatásban R5.7. miatt minden szakma számára szükséges a képzésekbe beépítve és továbbképzésként is az MI-projektekben való részvétel megalapozása (V.3.4.).

R5.11. Egy besorolás-mátrix modell segítségével (V.4.1.) a különféle rendszerek besorolása, vagyis azok szabályozásának alapja, áttekinthetőbbé és látványosabbá tehető.

VI. AZ ERŐÉRVÉNYESÍTÉS ÚJ KORSZAKA ÉS AZ MI

„*Az emberiség történetének eddigi legnagyobb geopolitikai forradalma az a digitális forradalom, amelyet többek között az MI terjedése tesz lehetővé.*"[2, pp. 24] Ennek a geopolitikai-digitális paradigmaváltásnak a széleskörű vizsgálata a kortárs hadtudomány egyik igen jelentős kihívása, ebbe kapcsolódik bele ez a dolgozat – főleg az alábbi fejezete révén. A korábbi részkutatásaim védelmi problémára (P2-re) irányuló szálai itt végre összefutnak, hogy elérjem a C2 célt[204]. Ehhez azonban K6 kutatási kérdéshez[205] úgy kellett a válaszok irányát kijelölni, hogy a vizsgálatok egyben H2 hipotézist[206] bizonyítsák, tehát ezt a két elvárást össze kellett egyeztetni. Mivel a C2 jóval szélesebb spektrumot ölel át, mint a H2, ezért ehhez igazodik az alfejezetek kialakítása, a bizonyításhoz szükséges részkövetkeztetések alátámasztása ezeken belül kap helyet. Nézzük a két elvárást külön. A C2 eléréséhez:

1. Az első alfejezetben általánosságban mutatom be, hogy a hatalmi erők érvényesítésének irányai az MI nélkül is változnak, ám ezen változásokhoz jól illeszkednek az MI képességei, sőt fel is erősítik ezeket a változásokat (VI.1.).
2. Ezután a hangsúlyt annak vizsgálatára helyezem át, hogy hogyan erősítik egymást ebben a vonatkozásban az MI és a kibertér technológiái (VI.2.).
3. Majd az első téma egyik konkrét aspektusát, a digitális térben kibontakozó visszaélési lehetőségeket elemzem, de az elterjedtnél jóval tágabb nézőpontokból (VI.3.).
4. Végül ismertetem a hipotézis két következményét: az MI hatására megváltozó katonai informatikát, és annak oktatási oldalát (VI.4.).

A H2 bizonyítását négy irányból (a fenti témákba ágyazva[207]) terveztem végrehajtani:

(1) Az erőérvényesítésben megjelenő paradigmaváltások területe felől, melynek bemutatása teljesen jelen fejezet feladata.

(2) Az autonómiakutatások és az MI-torzítások irányából. Ehhez itt le kell zárni a korábban beindított gondolatmeneteket, melyek az autonómiával kapcsolatos kutatásban, majd annak az MI kihívásokkal kapcsolatos folytatásában (a III-IV. fejezetben) vetődtek fel.

[204] C2: A világ változásának, valamint az erőérvényesítés új tendenciáinak MI-vel való kapcsolatát elemezve határozni meg az MI védelmi szempontból fontos tényezőit.

[205] K6: Védelmi kérdéskör: *Hogyan (milyen modellek, fogalmak, összevetések, elemzések mentén) ragadhatóak meg azok az újdonságok, melyek az MI hatására a védelmi szegmensben várhatóak?*

[206] H2: Az MI tovább fokozza és támogatja azt a tendenciát, melynek során az erőérvényesítésben folytatódik hangsúlyeltolódás a puha műveletek felé.

[207] Kivéve VI.4.-et, amely nem a bizonyításra, hanem az eredmény felhasználási lehetőségeire irányul.

(3) Az MI és a társadalom kölcsönhatásának irányából. Ennek vizsgálata korábban kutatásaim tárgyát képezte, ezeknek csupán legfontosabb részei kerültek ebbe a fejezetbe, azok, amelyek ide vonatkozó érveket is szolgáltatnak.
(4) Az első fejezetben bemutatott affektív számítástechnika (gépi érzelmek) felől. Vagyis a nem-intellektuális MI védelmi használatának elemzéséből is számítok érvekre.

VI.1. AZ MI ÚJ ERŐÉRVÉNYESÍTÉSI MÓDOK ESZKÖZE ÉS KATALIZÁLÓJA

A digitalizáció kínálta lehetőségek a korábbiaknál sokkal összetettebb paradigmaváltást hoztak az erőérvényesítésben. Itt azonban azt szeretném középpontba helyezni, hogy nem a technológia hajtja előre ezt a folyamatot, a technológia csupán jól kihasználható a világ adott változásának kiaknázásához. Ehhez olyan modelleket és terminológiákat dolgoztam ki, melyek a folyamatot ebben a komplex megközelítésben kezelik.

VI.1.1. Paradigmaváltás az erőérvényesítésben – a hadtudomány össztudományivá válása

A technológia hat a világra, de nem irányítja, hanem inkább katalizálja azokat a változásokat, melyekre megérett az idő. Erre felhozható, hogy számos találmány csak jóval később válik népszerűvé az egyéb összeadódó tényezők miatt. Például az elektromos autók több mint száz évvel megjelenésük után most kerültek hirtelen piedesztálra, és e mögött – legalábbis hivatalosan – olyan eszmék állnak (fenntarthatóság, környezetvédelem), melyek egy évszázada még nem merültek fel. A technológia erőltetett terjesztése miatt persze egyéb geopolitikai és gazdasági indokok, illetve a kőolajkészlettel kapcsolatos hosszú távú kérdések sejlenek fel – ám ezek is az ismertetett elvet támasztják alá, hogy a technológia nem irányítja a világot, inkább csak támogat bizonyos folyamatokat. Ezen elv alapján lesz érdemes elhelyezni az MI-t abban a paradigmaváltozásban, amely már akkor elkezdődött, amikor a neurális hálók még csak laborokban csíráztak.

A gyors változással hadtudományi szempontból nagy kihívás lépést tartani, de ez fontos feladat, hiszen a technológia fejlődése és a hadtudomány mindig szorosan összefüggött. Sok megközelítés szerint a lőpor kezdte el tudománnyá változtatni a középkori hadművészetet. Ám a társadalmi változások, az erkölcsi változások, a gazdasági erő átalakulása és a technológiában meginduló változások együtt alakították a világot, és annak hadi szegmensét is. A világképlet összetevői egymást erősítették – és ez most is így zajlik

Ezen komplex változás összetevői közül szempontunkból most pár dolog emelendő ki. Az egyik, hogy az országhatárok jelentősége lecsökkent, a gazdasággal együtt sok minden globálissá vált. A második, hogy az individualizmussal együtt felerősödött a jogi szemlélet, mely a nyílt agresszió minden formáját ellenzi, legyen szó nevelésben alkalmazott fizikai fenyítésről, a vitás helyzetek ököljoggal rendezésén át a családon belüli lelki terrorig. Ez a megközelítés pacifizmusként vagy legalább a nyílt agresszió kerüléseként is megjelenik, mely hat az országok erőérvényesítő politikájára is.

Az utóbbihoz kapcsolódik a puha műveletek 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.[208] Ezeket a hadtudományban együtt kezelik a kemény (fegyveres) katonai műveletekkel, vagy azok lehetőségével. Eszerint az úgynevezett hibrid műveletekben [244, pp. 18, 25] összehangoltan vetnek be a kemény módszerek támogatására puha műveleteket. Ezért a fejlődés jelen fázisában a rombolásra és emberéletek kioltására alapuló klasszikus katonai módszerek alkalmazása mellett az erőérvényesítés más módjai egyre inkább előtérbe kerülnek. Az így elvárt eredmények pedig jóval optimálisabban elérhetőek a számítástechnika, illetve az MIKT (II.1.) által kínált lehetőségek kiaknázásával.

A tudományok konvergenciája nem csupán az MI-t érinti (V.3.4.), de a hadtudomány, illetve az erőérvényesítés területén is jellemzővé vált. A politikai vagy gazdasági célok eléréséhez minden tudomány eredménye bevethetővé vált, így az erő kifejtése össztudományi perspektívát vett fel. A legkorszerűbb szervezési, kommunikációs, pszichológiai, szociológiai, gazdasági és társadalomtudományi módszerek legalább olyan fontosak, mint a high-tech eszközök, melyek a fizika, biológia, vegyészet, informatika és egyéb kutatások eredményeit hasznosítják. Ebben az összefonódási folyamatban egymást segítik a technológia és a több ezer éve ismert erőérvényesítő módszerek [245], így teljesen új perspektívák nyílnak meg.

VI.1.2. A virtuális erőközpontok modellje

A puha módszereket azonban nem csupán egy katonai jellegű művelet részeként lehet használni, ezért ezeket katonai alkalmazásuktól különválasztva is érdemes vizsgálni. Ezt indokolja a hátországok eltűnésének folyamata is a mai információs-, kiber- és gazdasági térben. Hiszen

[208] Megemlítendő, hogy az itt tárgyaltak a „közepes művelet” kategóriába tartoznak egy pontosabb, de nemzetközileg nem közismert, három szintű felosztás szerint. Ld. [243, pp. 23–24].

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 (I.4.). Ezért a segítségével végrehajtott befolyásolás által bizonyos célok hosszú távon eredményesebben elérhetőek lehetnek, mint katonai műveletekkel. Az így fejlődő digitális 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 államival, vagy egy független csoport, egy kis cég, mely aszimmetrikus puha műveletekkel[209] gyakorol nyomást. Úgy vélem, hogy a hatalmi struktúra ilyetén 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?

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 hajlandóak az egészségügyi szervezetek számára átadni a járvány jobb kezeléséhez szükséges adatokat. Erre reakcióként volt, aki ezeket a vállalatokat vádolta a digitális erejükkel való visszaéléssel: „*maguknak (ti. a vállalatoknak) van egy magánkormánya, amely döntéseket hoz a társadalma felett, ahelyett, hogy a demokratikus kormányok meghozhatnák ezeket a döntéseket*."[210] Más szerző viszont pont az állam (a regnáló

[209] 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.

[210] (saját fordítás). Ez a megfogalmazás a vállalatokat úgy azonosítja, mint amit mi a 2. fejezetben virtuális erőközpontokként említettünk. Ld. [246].

kormány) túlkapásaként, és a polgárok jogának megsértéseként értékeli az állam igényét az egészségügyi adatokra.[247] 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. A helyzet változásához jól illenek az MI képességei

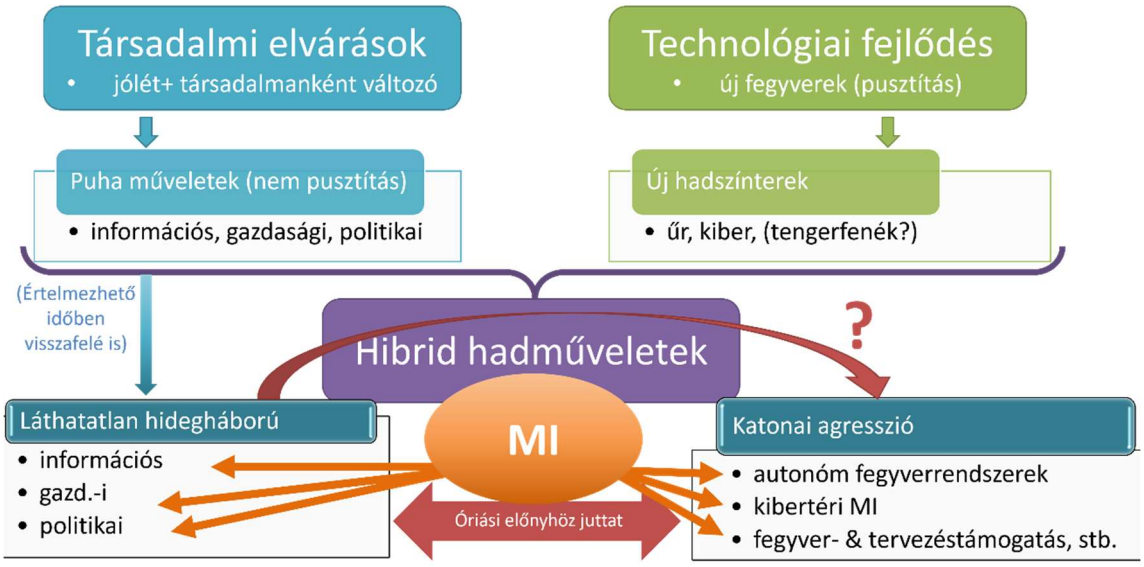
Folytatom a gondolatsort a virtuális erőközpontok modelljétől azzal, hogy pontosítom az MI és az új paradigmák kapcsolatát. Elindult egy folyamat, melyen látható, hogy a klasszikus hadműveletek céljait[211] már jócskán kitágította. Nem csupán a hadszínterek számának növekedéséről van szó, – bár igen lényeges változás, hogy az űr és a kibertér által szétmállik 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 fordítani.[212] Ezek az új hadszínterek azonban még nem implikálnák 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 céljaikat emberélet elvétele vagy látványos pusztítás nélkül, láthatatlanul érhetik 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 agresszió mentesség, és az államok között a nem fizikai, vagyis „puha hadviselés" formái kerültek előtérbe. Ez kihat az „igazságos háború" koncepcióra is [250], 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 [251], amely szerint a hibrid hadviselés fogalmába sorolandóak olyan célok is, melyeket a támadó fegyveres konfliktus nélkül

[211] Hagyományosan a térben körülhatárolható és tűzerővel kezelhető célokban gondolkodtak. ld. [248, pp. 372]
[212] További kibertéri lehetőségek: [249].

(vagy azon kívül) akar elérni. A nyomásgyakorlás és ennek ellenmódszere egyaránt lehet láthatatlan. Nem csupán a közismert kibertámadás és kibervédelem rejtett módjára kell gondolni,



17. ábra: A láthatatlan hidegháború az MI vonatkozásában (saját készítés)

hanem akár a kognitív befolyásolás [252] kifinomult módszereire az információs műveletekben, vagy gazdasági nyomásgyakorlásra, – de a diplomáciai kényszerítés eszköztára is ide tartozik.[253, pp. 22–28] Ezekhez a műveletekhez az MI olyan jelentős segítséget képes nyújtani, amely talán a rettegett harctéri használati módjain is túlmutat (ld. 17. ábra), hiszen, amint az már 1995-ben előre látható volt, a mesterséges (virtuális-, kiber-) térben a mesterséges elme sokkal otthonosabban mozog, mint az ember.[254]

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, többek között sok ilyen témájú cikk és hír generálásával. Vagy másik gazdasági példa: a ráhatások egyik formájában az ellenfelet az ellátási lánc nem feltűnő szintjein, apró alkatrészeken keresztül teszi sebezhetővé[213] vagy bojkottálja. Az MI otthonossága a kibertérben és alkalmazhatósága kognitív feladatokra akár egy COTS-rendszert[214] (ld. VI.1.4.) is védelmi képességekkel láthat el, például gazdasági manipulációra civil rendszerek is készülhetnek, hiszen a cégek egymást is támadják (ld. alább), ennek mintázatait az állami

[213] Például amikor kiderül, hogy az F-35-ös repülőgépekhez egy kínai tulajdonú cég gyárt kulcsfontosságú áramköröket. ld. [255].

[214] Boltban a polcról levehető (*Commercial off the Shelf*) termék.

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 az MI-vel összefüggő használati módokat 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.[256] Az Internet Research Agency emberi alkalmazottai még „csupán" az ellenfél belpolitikájába próbáltak beleszólni, [257] ám egy gépi elme sokkal átfogóbb károkozásra lehet képes.

A második világháború utáni korszakot máig főleg az elrettentő erő „hideg" növelésével, vagyis az 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 – ám ezt itt nem vizsgálhatom.

VI.1.4. Az MI direkt és az adaptált védelmi felhasználása

Mielőtt továbblépnék a gondolatmenetben, érdemes egy kis fogalmi kitérőt tenni. Meg kell ugyanis különböztetni két markánsan eltérő részt a védelmi informatika (és az MI) területén. 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*-t. Kiterjeszthető a felosztás a többi védelmi területre is, és mindenhova, ahol sajátos technológiákat használnak.

Napjainkig a katonai informatikát az jellemezte, hogy a civil életben terjedő technológiákat adaptálta katonai célokra. Igazából nem is volt szükség katonai processzorra vagy memóriára, illetve katonai operációs rendszerre. Elegendő volt a meglévő alkatrészeket egy olyan házba építeni, mely védi azokat a terep- és harci körülmények viszontagságaitól, és a szoftvereket bevizsgálni, 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 + portálnyelv + 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 egészségügyi diagnosztikai eszközt a civilek is hatékonyan használhatnak, ennek a harctéri verzióját

csupán jobban feltanítják a tipikusan katonai (lövés, repesz, robbanás stb. által okozott) sérülésekre, illetve a vegyi támadások pontos felismerésének nagyobb palettáját kapja. Szorosan ehhez kapcsolódik a COTS-eszközök, vagyis „a boltban a polcról levehető" (*Commercial off the Shelf*) konkrét árucikkek köre is. Egy 25 éve létező szabvány alapján sorolhatóak ide egyes termékek [258], 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 megjelenhettek a sajátos, kifejezetten katonai célú informatikai rendszerek. Persze, ahogyan egy nehéz harckocsiban is számos alkatrész és megoldás a teherautókéval azonos, úgy ezeknek is vannak civil termékekből örökölt alapjai. De ahogyan a harckocsira, 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 drónraj 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 rákötik 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,[215] 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, amellett, hogy a hagyományos kiberbiztonsági 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 harckocsi. Ez egyébként a katonai informatika területére is jelentős visszahatással van (ld. VI.4.1.).

[215] Egy vezetésirányítási kommunikációs számítógéprendszerhez (C4: Command, Control, Communication, and Computers).

VI.1.5. A mai és a klasszikus hidegháború összevetése

Miért is nevezhető a napjainkra kialakult helyzet egy második hidegháborúnak? 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 hosszú távú és tervezett befolyásolására. Az államok közötti erőérvényesítés ilyen nem fegyveres formáit hagyományosan nem tekintették harcnak, – a helyzet azonban változik. Először megmutatom a mai helyzet eltéréseit a II. világháború utánihoz képest, majd egy példával erősítem meg, hogy az eltérés igen markáns.

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 [259], 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 amellett, hogy mind a nyersanyagok, mind a piac szempontjából jelentősek ezek a területek, a multinacionális ellátási láncok is egyre kevésbé képesek 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. Itt jutottunk el oda, hogy feltegyük a másik kérdést: mi a szerepe az új hidegháborúban az MI-nek? A válasz az a párhuzam, 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[216]-ra irányuló innovációktól, számos cég és kisebb állam is hivatalosan elzárkózik [260], ám ilyen elzárkózást nem látunk sem a puha műveleteket, sem a hadászati tervezést vagy a művelet végrehajtását sokféleképpen támogató (pl. vezetéstámogató) [261] MI-vel kapcsolatban.

[216] Lethal Autonomous Weapon System, Halálos/Gyilkos Autonóm Fegyverrendszerek.

Fontos különbség, hogy a civil cégek a világháború után még nem rendelkeztek olyan hatalmi potenciállal, mint ma. Gazdasági erő szempontjából már jó ideje sokuk ereje meghaladja egy-egy országét. Ilyen cég is hajthat 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. Vagyis egy jelentéktelen cég is jelentős lépéseket tehet egy aszimmetrikus információs műveletben. Még nem látható annak pontos következménye sem, hogy a jelentős informatikai innovációk általában 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 célrairányíthatóság, mely a tömegpusztító fegyver mennyiségi szemléletével szemben technikai alapot ad az ellenség kulcsrendszereinek kiiktatásához, célra irányított kiberműveleten keresztül (konkrét példák VI.2.1.-ben), vagy kulcsszemélyiségek ellen személyre szabott dróntámadásokkal.[262] De a hátrányokozás mellett óriási előnyhöz is juthat például egy olyan gazdálkodás is, amely az MI meghatványozott elemző és szimulációs képességeire építi a célratörő döntéselőkészítést és a stratégiaalkotást.

Ez az újszerű hidegháború az ellenőrizhetőség terén is eltér a korábbitól. Még ha sikerülne is valami nemzetközi konszenzust kialakítani, sokkal nehezebb, szinte képtelenség lesz azt úgy ellenőrizni, mint egy atomleszerelési egyezményt. Egy informatikai kutatólabort álcázni nagyon könnyű, ráadásul egy MI-ben elrejtett veszélyes tudást pusztán teszteléssel megtalálni nagyon nehéz (IV.2.). Pl. az ún. generatív intelligencia (II.2.3.) alkotóképessége információs műveletekben már ma is komoly gondokat okozhat az álhírek gyártásától a hamisított képeken át ezek felismeréséig. Már működik pl. a technológia kockázatelemzésben való felhasználása [263], így katonai alkalmazása kikövetkeztethető [264]: alkalmazható újszerű hadászati támadó manőverek kidolgozásától az ismert hadi stratégiák elleni újszerű, hatékonyabb védekezési stratégiák generálásáig. Az 5. sz. táblázatban összefoglaltam a főbb különbségeket (az érzelmet később vizsgálom VI.2.3.).

Lényeges különbség továbbá az a probléma, hogy az állami, vállalati, terrorista (stb.) akaratokon túl megjelenhet egy újszerű kiszámíthatatlanság. Mind az MI-vel támogatott, mind a robusztus számítástechnikai rendszerek sajátos hibázási módokkal rendelkezhetnek. Ezért végeztem el a III. fejezetben a „gépi autonómia” részletes vizsgálatát, melyre hamarosan visszatérek (VI.2.4.). Itt elegendő annyit említeni, hogy egy felfokozott helyzetben valóban előfordulhat,

hogy egy gépi döntéstámogatás harcol egy másik döntéstámogató MI-vel, emberileg követhetetlen sebességgel. Ez pedig könnyen vezethet hibás döntések sorához. Itt újfent megjegyzendő, hogy ez nem csupán a katonai rendszereknél léphet fel, hanem pl. a gazdaságtámogató rendszerek vagy közvélemény-elemző és információstratégia-menedzselő rendszerek esetében is.

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**
az emberi akarat dönt	**megjelennek a túlautomatizált rendszerek döntési hibái**

5. táblázat: A XX. és a XXI. századi hidegháborúk főbb különbségei (saját készítés)

VI.2. AZ MI ÉS A KIBERTÉR HATÁSAINAK ÖSSZEADÓDÁSA A PUHA MŰVELETEKBEN

Ebben a részben néhány példán szeretném illusztrálni azt az állítást, – mely szinte axiómaként is kezelhető, – hogy az MI a virtuális térben van igazán elemében, ott használható a leghatékonyabban. Ez egyben azt is jelenti, hogy a digitális tér és az MI segítségével a korábbinál sokkal hatékonyabban, tisztábban, politikailag korrektebben, tehát az elvárásoknak megfelelőbben lehet kivitelezni az erőérvényesítés új formáit. Ezek közül vázolom az MI-alapú támadásokat és MI-t támadó módszereket, és az affektív számítástechnika védelmi lehetőségeit. Itt térek vissza az autonómia témájára is, mely ehelyütt már konkrétabban vizsgálható a vázolt védelmi perspektívában.

VI.2.1. Kibertámadások az MI segítségével

Korábbi tanulmányaimban az MI támadó célú használatáról általánosabb áttekintést adtam [265], és az alábbi vírusok alaposabb bemutatása mellett más modelleket is vázoltam [262]. Itt elsőként az IBM Research által kifejlesztette DeepLockerről ejtek szót. Ezzel demonstrálták a 2018-as Black Hat (USA) konferencián azt [266], hogy MI segítségével megkerülhetik az általánosan alkalmazott védekezési lehetőségeket. Laborkörülmények között létrehozott kártevőjük könnyen letölthető MI-modelleket kombinált ismert malware[217]-technikákkal. A cél az volt, hogy felkészülhessünk hasonló támadási helyzetekre.

Az IBM laborkártevője kétféle módon is kiaknázza az MI erejét: (1) a cél azonosításához és (2) a rejtőzködéshez. (1) A vírus olyan MI-triggereket (eseményindítókat) használ, mint például az arcfelismerés és a hangfelismerés (természetesen képes a hagyományos hely, rendszertípus, illetve konkrét eszköz valamely azonosítója stb. adatok alapján is aktiválódni). (2) Fő újdonsága a kitérés (elrejtőzés) eddiginél hatékonyabb módszerében van.

A kitérő technikák olyan módszerek, amelyeket a számítógépes támadások használnak a rosszindulatú tevékenységek elrejtésére. A támadások hagyományosan különböző kitérő technikákat használnak a biztonsági védelmi rétegek elkerülésére. A támadó alkalmazhat egy vagy több kitérő technikát. Már az 1980-as évek vírusainál elkezdték használni a kitérő technikákat. Az 1990-es évektől jelent meg a kód kártékony részének titkosítása, ami lehetetlenné tette a kódrészletek keresésének hagyományos, összehasonlításos módszerét. Ez ellen hozták létre a biztonsági oldalon a virtuális tesztkörnyezeteket, a *sandboxokat*,[218] A 2000-es évekre a vírusok már képessé váltak érzékelni, hogy speciális virtuális környezetben (homokozóban) futnak-e, vagy éles rendszeren. Ez ellen homokozó helyett úgynevezett „csupasz fém" környezetben tesztelnek.[267] Így a támadók újabb stratégia felé hajlanak: a támadás kifinomultabb *célzásával* érik el, hogy kártevőjük rejtve maradjon. A rosszindulatú kód csak akkor töltődik le, vagy csomagolódik ki, tehát csak akkor hajtódik az végre, ha a célpontot „tisztának" találja.[219] Csakhogy

[217] A malware a *malicious software* rövidítése, magyarul rosszindulatú számítógépes program.

[218] Ez a „homokozó" általában egy virtuális számítógép, ahol következmény nélkül el lehet érni, hogy a kártevő aktiválódjon, így igazi rendszerünkben nem képes kárt okozni, de vizsgálhatóvá válik.

[219] Ennek a „mesterlövész" módszernek egy korai (2010-es), igen hírhedett példája a Stuxnet féreg, amelyet úgy programoztak, hogy csak egy adott gyártótól származó specifikus hardver- és szoftverkonfiguráció (adott esetében iráni urándúsítók) esetén aktiválódjon.

ehhez is szükséges valami trigger – ezért a védekező oldal pedig ezentúl ennek felismerésére koncentrált.[220]

A DeepLockernél azért nem lehet az aktivációt kiváltó trigger-t megtalálni, mert az is titkosítva van, mégpedig egy mélyneurális hálózat alkalmazásával. De nem csak azt fedi el ez az MI, hanem a károkozó kódot, a konkrét célt, és az aktiválódás típusát (arc? hang? hely? rendszer?) egyaránt. A háromrétegű álca alatt csak akkor állítja elő az „indítókulcsot" (triggert), amikor minden elvárt körülmény együtt áll – ám akkor már késő detektálni. A kutatócsoport a fentiek demonstrálására az ismert WannaCry vírust (zsarolóvírust) álcázta a DeepLockerrel, egy jóindulatú videókonferencia-alkalmazásba rejtve el a kártevőt. Kiváltó körülményként az MI-modellt úgy képezték ki, hogy felismerje egy adott személy arcát, és csak ennek hatására bontsa ki és indítsa be a vírust.

A másik fontos modell a rajintelligencián alapul. Ennek korábbi bemutatása kapcsán (II.3.3.) említettem, hogy egy biológiai „raj" számítógépes implementációja egy teljesen elosztott rendszer is lehet. Ahogyan az élő rajközösségek életképességét nagyban emeli a rajtudás, úgy a virtuális térben is nagy előny ez a képesség, főleg a központi vezérlésű informatikai módszerekkel szemben. Amikor ezt az elvet kibertámadásokra használják, akkor a környezetükhöz alkalmazkodó víruskódok jönnek létre. A különféle hátráltató tényezőket a vírusraj entitásai közlik egymással, ez alapján új, javított támadó példányok születnek: az ehhez szükséges módosításokat hozza létre az újszerű tanulórendszer. Ráadásul ez a módszer képes a rajintelligencia közös tudását felhasználni a kártékony kódok elrejtéséhez (mimikrijéhez) is. Tehát a támadó kód, mint egy kaméleon, az adott környezetnek megfelelő legoptimálisabb technikát alakítja ki az ellen, hogy megtalálják. Ezek alapján már világos, hogy egy ilyen vírusrajjal hatékonyan lehet tömeges (robusztus) kárt okozni egy vagy sok rendszerben, hiszen:

- a rajnak nincs központi vezérlőegysége;
- a rajnak lehet kollektív emlékezete, megoszthatja tudását valamilyen stratégiáról (sikeres/sikertelen) és „tanulhat belőle";
- így a raj egyfajta közösségi és kommunikációs hálózatot hoz létre. Minden tag különféle módon kommunikálhat másokkal. (Az információ átadódhat közvetlenül, kiválasztott egyéneken keresztül, egyénről alpopulációra stb.);
- a raj alkalmazkodási képességét rejtőzködésre lehet használni.

[220] Vagyis a védelmi program automatikusan rákeres a „ha ez történik, akkor hajtsd végre ezt" típusú kódsorokra, azt megtalálva a rendszer jelez, a szakemberek pedig megtalálhatják a probléma forrását.

Tehát ha nagy nehezen észre is vesz a védelmi rendszer néhány víruspéldányt, azokat hiába távolítja el, mert az nem érinti a teljes raj működését. Nyilván egy ilyen támadás korábbi más, már bevált rosszindulatú technológiák (férgek, malware-ek) módosításával készül, azok ártó tulajdonságait nagyíthatja fel óriási mértékben.

Egy cseh kutatócsapat az osztravai egyetemen Ivan Zelinka vezetésével publikált is egy olyan vírusstruktúrát, amely ilyen elven működik.[121, pp. 207–224] Laboratóriumukban kísérleti mintát hoztak létre egy raj-malware megvalósítására. Konkrétan egy „klasszikus" botnet[221] vírusból indultak ki. Ezt alakították át egy önjavító-önreplikáló kártevőstruktúrára. A célhoz három technológiát ötvöztek: a számítógépes vírus alapelveit, a raj intelligenciáját és a komplex hálózati elemzési képességet.

Ezek után nézzünk példát arra is, amikor MI-vel, de nem vírus által jön létre a támadás. A szélhámosságot (kibertéri nevén *social engineering*-et) például az MI rendkívül jól támogatja, épp embert utánzó alapjellege miatt. Én kétfelé osztom ezt a digitális csalásformát.[222] Az MI-használó social engineering egyik része a valóság digitális utánzásán alapul, nevezzük ezt „utánzó szélhámosság"-nak. Ide tartoznak a deepfake technikák (MI-technológiák segítségével meghamisított videók), vagy a digitális megszemélyesítés. Régi példa erre az az eset, amikor egy bankvezető hangját és beszédmódját a banki beosztottaknak mesterségesen utánozva csaltak ki átutalásokat a támadók.[269] Az MI által könnyen készíthető képek, szövegek, zenék és videók ma már igen könnyen használhatóak információs műveletekben is.[223] Az MI alapú social engineering másik részénél az emberek személyes, egyedi szokásrendszerének egyre tökéletesebb feltérképezése a kulcs, nevezzük ezt „kutató szélhámosság"-nak. Ez hasonló ahhoz, amit a tudósok „a social engineering információgyűjtő fázisa"-ként említenek.[270]. Ezzel érzékeltettem a problémát, ezért a MI-alapú támadások rendkívül sokrétű további lehetőségei közül csupán egyet tartok fontosnak majd még megemlíteni (VI.2.3. végén).

VI.2.2. Az MI kiaknázása támadásra és védelemre

A IV.2. alfejezet elemzését nincs is szükség túlmagyarázni, hiszen szinte kézenfekvő, hogy az ott vázolt torzítási komponensek egyben sérülékenységek is. Ott a megközelítés a tudattalan hibaokozásra koncentrált, – de mindegyik megjelölt aspektus tudatosan is használható egy MI ágensekkel felszerelt rendszer támadására is.

[221] Botnet – robot network. A támadó vírussal fertőzött számítógépek seregét használva támad.

[222] A szakirodalom egyéb felosztásaitól (pl. [268]) ebben eltérek.

[223] Léteznek műfajok, melyekhez nem kell jó minőségű videó: pl. egy lejáratásnál elegendő a szatirikus utalás.

Ebből a szempontból is kiemelendő az emberi aspektus. Az ember az MI szempontjából is a „leggyengébb láncszem", akár a fejlesztői vagy tanítói hibákat, akár a felhasználói hibákat nézzük. Hiszen az emberi tényező lesz az, aki létrehoz valamilyen a sérülékenységet a rendszerben (akár tévedésnek álcázva), a hat bemutatott szegmens egyikében. Hasonlóan ahhoz, ahogyan szándékosan hagyható hátsó kapu egy hagyományos rendszerben (elrejtve), úgy ilyen jellegű tervezetten torzító rendszer sem elképzelhetetlen. Tehát az algoritmus, illetve a modell szintjén is lehet támadni, pl. úgy, hogy a szabotőr egy kicsit hibásan programozza le a modellt, vagy netán a matematikai modellbe csempészik hibát. Vagy csak gyengébb modellt készít, mint a riválisok: hiszen a kibertérben kibontakozóban van „a matematikák összecsapása", ahol az a kérdés, hogy melyik fél matematikája a lehető legerősebb.[224] Azonban egyelőre jóval egyszerűbb, így jóval valószínűbb olyan jellegű szabotázs, mely az MI-hez társuló hagyományosan programozott részen hozza létre a hagyományos sérülékenységet, vagy akár a fejlesztő által választott hardverelemben van előre beépített sérülékenység.

Ezeknél azonban lényegesebben ismertebb és jobban dokumentált az adatmérgezés típusú támadástípus, mely a tanítóadatokat, illetve a tanítási metódusokat érinti.[271] Egy jól megtervezett tanulás-fertőzéssel elérhető, hogy a rendszer néhány irányzottan rossz választ hozzon, a támadó szándéka szerint manipulálva az MI válaszait vagy döntéseit. A támadástípus inkább az MIKT-rendszereket érinti, hiszen hagyományos védelemmel is ellátott zárt fejlesztések esetén csak szabotőr(ök) által megvalósítható. MetaPoison néven erre is mutattak be demonstrációs fertőzést [272], melyet az akkori Google-felhő egyik API-ján teszteltek. Ez egy olyan általános célú adatmérgezés-segítő keretrendszer, mely automatizálja a korábban inkább emberi erővel történő folyamatokat, és segítséget adhat egy támadónak akár finomhangolási, akár nulláról történő tanításkor végzett támadás esetén.

Szót kell még ejteni itt nagyon röviden néhány egyéb támadásmódról is. Az MI-vel kapcsolatos félelmek egyik része azokkal az adatokkal kapcsolatos, melyeket egy nagy MI-szolgáltatás a felhasználóktól megtanul. A neurális háló feketedoboz jellege miatt ezek ugyan nem úgy hozzáférhetőek, mint egy űrlapon megadott (esetleg bizalmas) adataink, viszont képet adnak a támadónak rólunk, felhasználhatja saját MI-rendszerének betanításához, de klasszikus kibermódszerekhez is alkalmazhatóak, például jelszótöréshez személyes adatokból tippeket kaphat belőle (erre célzott prompt-tal).[273]

[224] A szerző (Da Silva) szerint ez egyben egy MI fegyverkezési verseny [271] – ami hasonlít az én felvetésemre (VI.1.5.).

Itt említendő, hogy a fenti módszerek azért hatékonyak, mivel erős aszimmetria valósul meg bennük a támadó javára. A rendszerek komplexitásával óriásira nőtt digitális támadási felületeket nagyon nehéz védeni.[274] A támadó pásztázza a rengeteg lehetőséget, és találhat benne rést, legyen szó egy operációs rendszer vagy program sebezhetőségeiről, vagy akár egy hardverelem gyenge pontjairól. Ezzel összecseng az a véleményem, mely szerint a hagyományos és az MI-rendszerek biztonsági szempontból konvergálnak (IV.3.), hiszen ennek a hátterében is a rendszerek átláthatatlan robusztusságát jelöltem meg.

Végül itt csupán említés jelleggel szólok a védelmi területről, hiszen erről írtam bővebben [175], és alapos elemzése nem szükséges az itt megjelölt célhoz. Először is azt emelem ki, hogy a támadásra használható módszerek mindegyike alkalmazható védelemre is. A támadásoknál kihasznált széles digitális támadási felületeket lehet javítási céllal is szkennelni. A fentebb említett támadás-demonstrációk segítségével kidolgozhatóak az ellenmódszerek. Másrészről pl. az embert utánzó social engeneering támadásokat egy MI-alapú gyorselemzés könnyebben kiszúrja, mint a hagyományos program vagy egy fáradt felhasználó.

VI.2.3. A „gépi érzelmek” néhány védelmi alkalmazási lehetősége

Az affektív számítástechnika példájával jól demonstrálható volt az intelligenciafajták sokféleségének gépiesítése (I.4.), és érdemes a védelmi technológiák változására erről a területről is 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 ezzel a legjobban kutatható résszel: vagyis a vállalati szegmens felől.

A jelenleg zajló gépi érzelem (GÉ) fejlesztések alapján a technológia terjedése céges környezetben[225] előbb várható, mint a katonaiban. Erre az ügyfélkezelés érzelmet kifejező chatbotjaitól [276] 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 eltöltése során. Ez nyilván nem nevezhető sajátosan katonai vagy védelmi alkalmazásnak. Ám mielőtt megnézzük, milyen sajátos alkalmazási lehetőségek kínálkoznak a jövőben a GÉ képességeinek kiaknázására a fegyveres erők és az állami hivatalok számára, vessünk egy pillantást a katonai GÉ múltjára.

[225] Egy rövid, tömör felsorolást ad erről pl. [275].

Ahogyan az MI katonai alkalmazásának megjelenése után nem sokkal kutatások tárgya lett, úgy a GÉ katonai felhasználása is már akkor felmerült, amikor még kevéssé volt ismert ez a technológia. 2004-ben a DARPA már „*nem invazív érzelemfelismerő rendszer (...) kifejlesztésére szólít fel, amely alkalmas katonai / operatív környezetben vagy olyan környezetben történő telepítésre, amelyben az ellenség lehetséges fenyegetéseinek diszkrét megfigyelése kívánatos*”.[226] Ez az érdeklődés napjainkig kíséri a fejlesztéseket, megjelenik a NATO 2020-as fejlesztési tervében is [3, pp. 52, 58], de még ekkor is csak az érzelem-input kerül fókuszba. Érdekes módon az érzelem-output alkalmazhatóságának vizsgálatát nem említi a DARPA 2020-as költségvetése sem.[277]

Konkrét felhasználási módok az idézett forrásokban nem találhatóak, ezért állítottam össze az alábbi listát néhány példával, melyben többször utalok korábban (I.4.) leírt technológiákra. 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).

1. **A kiképzés területén** főleg a már meglévő MI-alapú, esetleg VR[227]-technológiával megoldott szimulációknak adott 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 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, mivel a szimulált arcok, emberi hangok, elhangzó megfogalmazások érzelmeket közvetítenek, ezért bizonyos pszichológiai műveletek kivédésére is hatékonyabb felkészülést tesz lehetővé. Végül megemlíthető, hogy akár a távoktatási feladatoknál is hasznos lehet, ha az oktató automatikus visszajelzést kaphat a GÉ-szenzor által arról, hogy a hallgatók mennyire követik az anyagot, mennyire tartják azt megfelelőnek.
2. **Az egészségügyi gyakorlatban** számos területen fontosak a tünetek kifejeződései (affektusai): az arckifejezések, az izomfeszültség, a testtartás, a kéz- és vállmozdulatok, a beszédminták, a szívverés, a légzés, a pupilla kitágulása, a testhőmérséklet.[40] Bevetés tá-

[226] 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: [50, pp. 96].
[227] Virtual Reality, azaz virtuális valóság.

mogatásához egy szemüvegbe integrált szenzor és például egy EPU-stick segítheti a diagnózist orvos nélküli helyzetben is, de ilyen helyzetekben egy testszenzoros rendszer elemző moduljaként is hatékonyan alkalmazható a GÉ. Emellett kifejezetten katonai lelki betegségek (poszttraumás stressz, szerzett fóbiák) kezelésére jó ideje alkalmazzák a hagyományos MI t – amely GÉ-vel bővítve pontosan érzékelni képes a páciens érzelmi reakcióját, vagy készülő pánikrohamát a kezelés során. Az ilyen kezelés is elsősorban a virtuális világok előállítására épül, ahol a rendszer reakcióit, amelyet korábban a kezelőorvos felügyelt, automatikusan meghatározza a GÉ, az előző pontban leírt szimulációhoz hasonlóan.

3. **Nemzetbiztonsági használatában** a csoportos hangulatmérés-elemzés során alkalmazható a GÉ. Hasznos lehet mind a saját állomány hangulatának (kimerültségének vagy lelkességének) elemzésére, mind pedig a civil lakosság reflexiómérésére. Ez utóbbira egy megvalósult példa az Arab Emírségek egy érdekes fejlesztése. Egy érzéselemző térfigyelő rendszer telepítésétől és használatától várják állampolgáraik hangulatának jobb megértését, melyek az állam döntéseinél figyelembe vehetőek.[41] (A jogi aggályok kezelése persze kérdéses.)
4. **Rendészeti területen** kihallgatások során egészen bizonyos, hogy tíz éven belül (beszerzési ártól függően) felválthatja a hagyományos hazugságvizsgáló berendezéseket a GÉ-alapú kikérdezéstámogatás. Akár a vizuális érzelemelemzési módszerek, vagy például az aRadar (ld. I.4.2.[228]) évről évre árnyaltabb képet ad a kihallgatott személyről (szorong, dühös, titkol valamit stb.). További előnye, hogy a vizsgálat rejtve is elvégezhető – miután leküzdötték ennek nyilvánvaló jogi problémáit.
5. **Titkosszolgálati vagy terrorelhárítási célú használatnál** az automatikus célpont-azonosítás eddiginél hatékonyabb megoldásait valósíthatja meg a GÉ. Például fokozottan biztosítandó helyszínen (fontos állami hivatal, bank), vagy kiemelt rendezvényeken segíthet kiszűrni 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 következtetni, amit ember nem vehet észre.

[228] Amint ott is említettem, a kutatásról információ – a fejlesztő cég beolvadása miatt – jelenleg ugyan nem elérhető, de valószínűtlen, hogy kidobták. Vagy titkos használata zajlik, vagy más néven kerül majd valamikor forgalomba.

6. **Döntéstámogató és döntéshozó rendszerekben** a GÉ alkalmazása elkerülhetetlen lesz, hiszen a várható emberi reakciók prognózisa nem csupán a puha műveletek során segít, hanem az állomány állapotának figyelembevételével a harci morál javításához is hozzá tud járulni.
7. **Információs és kiberműveletek** területét együtt említem – és azért utoljára, mivel ez a GÉ talán 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 humán biztonsági réseinek megtalálásához.[42] Hiszen közvetlenül a személyt elemezve, vizuális érzékeléssel lehet megkeresni a kulcspozícióban lévő emberek 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, pletykakeltés). Védelmi szempontból az ilyen manipulációk felismerésére jó.

VI.2.4. Az autonómia katonai és védelmi alapkérdései

Az atomkorszak fenyegetése nem múlt el, de az emberek ma már inkább az MI által vezérelt autonóm fegyverektől, a „gyilkos robotoktól” félnek. Ezzel szemben – mint látni fogjuk – a valóság az, hogy a haladás által nyújtott lehetőségekkel egyelőre nem a hagyományos katonai területen lehet leghatékonyabban visszaélni. A „gyilkos robot” kifejezés eleve félrevezető. Hiszen a *robot* szót lehet „eszköz” és „döntéshozó alany” egyaránt. Az előbbi esetben a „gyilkos kés”-sel állítható párhuzamba, a második a „gyilkos ember”-hez hasonlatos képzeteket szül. Sok kontextus ez utóbbit próbálja meg sugallni. Ebben a sugalmazásban azonban a disztópikus fantasy irodalom keveredik a valósággal, ezért szükséges a kérdést helyretenni. A tiltakozás és a félelem háttere tehát, hogy a mai technológia lehetőségeit keverik az elképzelt, saját tudattal rendelkező gépek mítoszával.

A fejezet eddigi elemzéseiben kimutattam, hogy az MI a puha műveletek területén rendkívül jól használható, nem tértem ki azonban a kemény műveletekben való bevetéseire, illetve azok

szunnyadó potenciáljaira. Ez utóbbi témával terjedelmes irodalom foglalkozik, melynek recenzióját sem kívánom itt nyújtani, csupán saját aspektusomra szorítkozom: összevetem az MI kemény és puha műveleti lehetőségeit. Ehhez az autonómiáról végzett vizsgálataim (III.) fonalát veszem fel, melyet az automatika és az autonómia hibázásával kapcsolatos elemzés végén (IV.3.) elkezdtem továbbgondolni – ezt a gondolatsort szeretném itt befejezni.

A gépi szabadság lehetőségeinek vizsgálata rámutatott, hogy a jelenlegi vékony MI valójában inkább egy fejlett automatika, melyre az önvezető autók hivatalos besorolása sem az autonómia szót használja (III.3.2.). Továbbá bizonyítottam, hogy egy fejlett erkölcsi érzékkel rendelkező ember szabadsága egy gépben nem utánozható le, egyrészt azért, mert annyira lényegi különbség van a gépi és az emberi autonómia között, másrészt mivel a világ leképezése nem lehet az ehhez szükséges mértékben objektív (III.5.). Ezeket figyelembe véve megállapítottam, hogy a jelenlegi vékony MI-megoldások tudatra ébredésétől nem kell tartani. Sőt, a mai rendszerek valójában nem képesek az emberi szándékkal ellentétesen sem tevékenykedni. A hírek, melyek időről időre ilyen esetekről tudósítanak, valójában inkább bulvárosítanak, nem pedig „tudós"-ítanak. A pontos dokumentációkat ugyanis nem adják közre, pedig azokból feltehetőleg világos lenne, hogy vagy az alkotók tudatosan tanították a gépet arra, hogy tehet olyasmit is, amit a netikett nem részesít előnyben, vagy pedig nagyon rosszul tervezték meg azt, hanyagul végezték el a munkájukat. Ezen a téren viszont nincs eltérés a hagyományos, programkóddal üzemelő gépektől, ahol épp így elképzelhető hanyagság vagy szabotázs (IV.3.).

Ezeket alkalmazva a fentebb bemutatott kétféle műveletre, világossá válik, hogy sem a kemény műveletek támogatására bevetett MI, sem a puha műveletekre alkalmazott neurális (vagy egyéb) tanuló rendszerek „elszabadulása" még jó ideig nem lehetséges olyan értelemben, mint ahogyan azt a disztópikus művészeti sci-fi vagy fantasy ábrázolások bemutatják. Emellett valóban nem minden innovációnál garantálható, hogy az MI etikájának hivatalos megközelítéseit és elvárt irányelveit[229] tartja szem előtt, és teljesíti a robotetika az „alulról" vagy „felülről"[230] fejlesztésével kapcsolatos elvárásokat. Valójában azonban sokkal könnyebb, hatékonyabb, olcsóbb és reálisabb lenne hagyományos, programozott módon valósítani meg valami gonosz világuralmi vagy világelpusztító tervet, – erre remélhetőleg a modellek bonyolultságának érzékeltetése (II.) elegendő érv. A disztópikus alkotások gyakran össze is mossák a hagyományos és MI-rendszereket (V.1.2.), vagy pedig úgy humanizálják a gépeket, mint ahogyan a kutyákba

[229] Erre számos példa található itt [278].

[230] Felülről tanítva szabályokat mondunk a gépnek, alulról a gép fedezi fel a szabályokat. E tekintetben egy átfogó EU-s jelentés: [279, pp. 90–91].

vagy más házi kedvencekbe vetít bele a gazdájuk számos emberi vonást, amely nem létezik. A jobb műalkotások a gépek elszabadulását az emberi hibákkal hozzák összefüggésbe, melynek fontosságát magam is hangsúlyoztam (IV.2.2.), – de mivel ezek a gépek elektromosak, ezért jóval könnyebben megfékezhetőek, mint például egy genetikai kutatás során történt hibázás.

Nézzük meg külön-külön is a kétféle művelettípust. A kemény műveletek során először is meg kell állapítani, hogy a valóságban olyan számú nem várt tényező lép fel, mellyel egyelőre egy legkomplexebb szenzorrendszerekkel támogatott fedélzeti számítógép még nem igazán képesek megbirkózni az MI erőforrásigénye miatt. De ennél súlyosabb az erkölcsi probléma, mely az autonóm fegyverrendszereket körbelengi, ennek megválaszolásához pedig az autonómia-vizsgálatok másik részére szükséges visszautalnom (III.4.1., mely III.2-n alapszik). Mert ha majd meg is birkóznak ezek a robotok a széllel, a látási viszonyokkal,az ellenséges zavarással, akkor sem úgy fognak ténykedni, mint egy vérszomjas állat, aki a gyilkolás örömeért öl. Akkor is közkatonaként fognak parancsot teljesíteni, ha a cél bemérését és likvidálását, valamint saját kitérő mozgásaikat maguk hajtják végre.[231] A kemény műveletet nem csak végrehajthatja, hanem támogathatja is az MI, ezért sokan attól félnek, hogy egy ilyen gép magától lő ki például rakétákat, melyek miatt a harcok elmérgesednek. Csakhogy a parancsnoki döntéstámogató rendszerek felépítésükből adódóan eliminálva vannak ettől, mivel nem létezik olyan helyzet, amikor előnyösebb, hogy egy ilyen súlyos, a saját országra is katasztrofális döntést (az ellentámadás miatt) a gépi szubjektivitás dönthessen el. Ezt minden sereg tehát saját érdekéből is az emberi (politikai, katonai felsővezetői) szubjektivitásra kell, hogy bízza.[232] Az az eset, amikor számítógépek a fizikai térben, emberre is ártalmas fegyverekkel harcolnak egymás ellen nem zárható ki egy távolabbi jövőben, – ám ehhez sem szükségesek MI-rendszerek; ugyanez a félelem az atomkorszakot is jellemezte már.

Ez nem jelenti azt, hogy az ilyen önvezető járművek nem veszélyesek olyan szempontból, hogy elromolhatnak, vagy megsérülve veszélyesen működnek. Csakhogy ez a hagyományos robotok esetében veszélyesebb. Hiszen ott lehetséges, hogy az irányító modult tartalmazó rész sérül, de a tüzelést szabályozó modul nem, így az eszköz folyamatosan tüzel, ha mozgást érzékel, és nem lehet leállítani. Ennek sokkal kisebb a veszélye egy neurálisan tanított gépnél, mivel, ha a mélytanuló sejtek egy része fizikálisan sérül (pl. mert a három GPU közül, melyen fut a rendszer, az egyik kiég), akkor könnyen az egész használhatatlanná válhat, összeomolhat,

[231] Vagy pl. az amerikai ACE koncepcióban működő légitámadás, melyben egyetlen pilótát autonóm eszközökből álló raj kísér, megsokszorozva ütőképességét.[280]

[232] Ilyen döntés nem lehet objektív, csak a „rendelkezésre álló információk alapján legjobb" döntést hozhat akár ember, akár gép, ezért tudományos értelemben való objektivitás nem merülhet fel.

hiszen a tanult válasz nem képes lefutni.[233] Ha nem fizikai sérülés történik, vagy csak kevés neuron sérül, akkor pedig a tanult képességekben (pl. alakfelismerésben) történik minőségi romlás, amelyet a rendszer felismerhet, és vagy javíthat vagy biztonsági módra válthat [282] [283]. Tehát valójában a hagyományosan programozott harctéri robotok veszélyesebbek, mint a neurális társaik.

A fentebb azonosított fontos tényezők: a nagy erőforrásigény, a fejlettség és az összefüggő neurális működés a puha műveleteket nem hátráltatja. Azért is van jobban elemében az MI a virtuális térben, mivel ott lépnek fel az eszköz mozgatását és tájékozódását nehezítő komponensek, mint a fizikai robotoknál. Másrészt, amint azt a fejezet eddigi részében kifejtettem, az információs, gazdasági és pszichológiai műveletekben, emberutánzó jellege révén, igen hatékonyan használható az MI és a GÉ. Harmadrészt ezek az alkalmazások nem a jövő disztópikus látomásai, hanem jelen vannak vagy lehetnek életünkben. És mint köztudott, egy veszélyes létesítmény (pl. a legtöbb magas rendelkezésre állású IT rendszerrel ellátott objektum) elleni kibertámadás hatása lehet tragikus a valós térben is, hiszen okozhatja ingóság, ingatlan vagy humánerő sérülését vagy elvesztését. Ezért állítom, hogy egy ideig még sokkal nagyobb veszélyt képvisel az MI a puha műveleti térben, amely ellen nincsenek nemzetközi tiltakozások és ENSZ nyilatkozat, mint a gyilkos robot nevű „mumus" ellen, amellyel korunk műszakilag kevésbé tájékozott tömegeit riogatják.

VI.3. ÉLNI ÉS VISSZAÉLNI A DIGITÁLIS TÉRBEN

Az előző alfejezetek tágabb elemzéseit itt a digitális visszaélésekre szűkítem. Olyan erőérvényesítési formákra koncentrálok, melyeket a számítógépes hálózatok tesznek lehetővé, melyek ereje – ezzel együtt a visszaélések kifinomultsága is – az MIKT által a jövőben várhatóan megtöbbszöröződik (II.1.). A digitális visszaélések elterjedt megközelítését a virtuális erőközpontok modelljére (VI.1.2.) alapozva lehet komplexebbé tenni. Ehhez a terminológiák tisztázása után a visszaéléseket a szokásosnál szélesebb perspektívában vizsgálom. Végül az orosz példán keresztül elemzem a témát.

[233] Megjegyzés: jóval a kutatás lezárása után, 2025 nyarán megjelent egy új, hierarchikus érvelési modell (*Hierarchical Reasoning Model*, HRM), mely hasonlóan alhálózatokból épül fel, mint az emberi agy, illetve a hagyományos IT rendszerek. [281] Ilyen HRM-ek esetén, pl. ha az alrendszerek fizikailag is elkülönülnek, akkor ezek részleges sérülésére nem vonatkozik ez az érv. Ám egyrészt a gondolatmenet e nélkül is igazolható, másrészt a HRM-ek valódi implementációnak terjedése esetén lesz a kérdés vizsgálható.

VI.3.1. Szuverenitás, ökoszisztéma és tekintélyelvűség a digitális térben

A digitális visszaélések témaköre két oldalról közelíthető meg: (I.) a *jogérvényesítés* felől, és (II.) a *jogsérelmek* felől. Az első terén inkább a gazdasági szereplők és az államok érdekei és igényei érdekesek, a másik szempont az emberi jogi szempontok alapján vizsgálódik. Alább először vázolom, hogy a szakirodalom fogalmai inkább az utóbbi megközelítéshez kapcsolódnak, ezután térek rá a két aspektus együttes vizsgálatára.

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*.”[284, pp. 6] 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.[285, pp. 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 [286]).

Másrészről nem csupán vállalatokra, hanem államokra is igaz, hogy a digitális szuverenitásnak a megfelelő színvonalú digitális ökoszisztéma szükséges, ámde nem elégséges feltétele. 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 [284, pp. 4], és hazánk sem élenjáró ebben, habár a probléma felszámolása folyamatban van.[234]

Rátérve a jogsérülés (II.) oldalára: ezt a kérdést a digitális tekintélyelvűség (*digital authoritarianism*) kifejezéssel közelítik (elsősorban amerikai jogvédők). Van, aki kifejezetten a „*tekintélyelvű kormányzás fokozására vagy lehetővé tételére szolgáló technológiák*" -at [287] érti alatta, de ezt inkább arra alkalmazzák, hogy egyes országok különféle antidemokratikus gyakorlatokat építenek ki digitális technológiák segítségével. Ez a megközelítés az állam biztonsága szempontjából abban a felismerésben válik érdekessé, hogy minden kutató igen baljósnak tartja saját (nyugati) országára nézve is egy másik állam belső (önnön 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.[235] 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.3.2. A digitális visszaélés irányai az erőközpont-modell alapján

A II. megközelítés szakirodalma alapján a digitális visszaélés három iránya bontakozik ki:

1. Állami szervek használhatják belföldön társadalmi csoportok és saját polgáraik ellen;
2. Átadják (exportálják) külföldre, és ott a külföldi állam az 1. pont szerint alkalmazza. Ezáltal ráadásul erősödik a kötelék is ezek között az államok között;
3. 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).[236] Véleményem

[234] A 2022-es helyzet alapos, önkritikus és előremutató elemzése: [285, pp. 32–104].

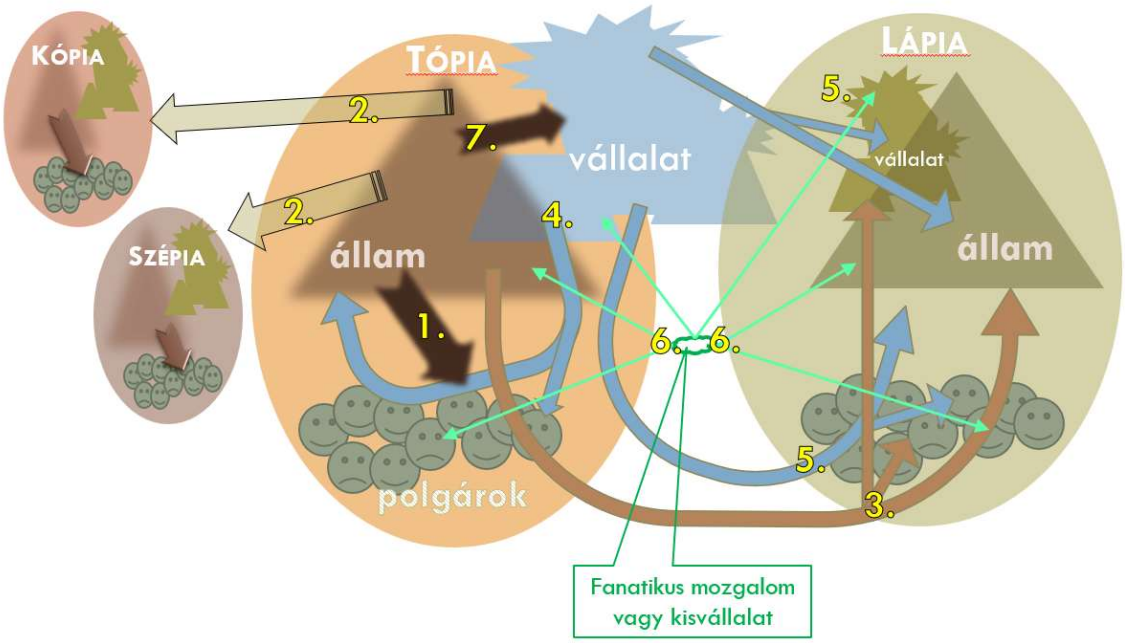
[235] ld. https://www.power3point0.org/about/ - ez a szervezet a jelenség dokumentálásával foglalkozik.

[236] Mely szerinte demokratikus országok belügyként kezelhető.[288, pp. 2]

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 értelmezhetők, sőt a nemzetközi szervezetek is kihasználhatóak. Így a digitális visszaélések a következő irányokkal egészülnek ki:

1. Erős vállalati erők „belföldön” vetik be, mozgalmakat dotálva vagy közvetlenül;
2. Erős vállalati erők „külföldön” az adott ország sajátosságainak megfelelően vetik be;
3. Független kis csoportok vetik be, általában aszimmetrikus műveletként;
4. Állami szervek saját vagy globális vállalat(ok) ellen vetik be;
5. (+ 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 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 tópiai 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.3.3. Az orosz modell elemzése

A fentiek szemléltetésére és a nyomásgyakorlás paradigmaváltására (VI.1.) az orosz állam digitális védelmi modelljét mutatom be, annak túlzásaival és indokaival együtt, mely önmagában is hasznos és aktuális információkkal szolgálhat. Ez a digitális tekintélyelvűség szakirodalmának egyik fő kutatási területe (a kínai modell mellett). A kínai és az orosz modellek közötti egyezés, hogy egyik sem a nyugati liberális demokrácia elvei szerint használja a digitális lehetőségeket, és mindkettő exportálja[237] az elért eredményeket (2-es irány). Az eltérés azonban jelentős. Míg a kínai egy impozáns, de költséges technológia, mely egy totális kontroll irányába fejleszt, addig az orosz modell inkább azokra a területekre koncentrál, amelyek olcsóbban biztosítanak digitális kontroll- és befolyásolási lehetőségeket.[290]

(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. Putyin elnök 2017-ben kijelentette, hogy az lesz a világ ura, aki az MI-szférában átveszi a vezetést, [291] 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 a vezető szerep megszerzése[238] (8-as irány).

A technológiai hátrányt azonban nem lehet csupán politikai súllyal kompenzálni, ezért már igen régen próbálnak a világpiacon MI-termékekkel is megjelenni.[239] A világelsőségtől távol vannak, de számos eredményről olvashatunk.[240] Bár folyamatosan emelkedik a szféra támogatása, még messze elmaradnak a kínai MI-szinttől. Kevés a kimagasló teljesítményű MI-alkalmazás (azért akad[241]), és tudományos áttörésekre úgy tűnik, nincs erejük. Ám az orosz digitális visszaélési modellhez ez is megfelel. Ezt mutatják a belföldi (1-es irány) példák – a Navalnij

[237] A posztszovjet országokon kívül a világ 26 országát említi.[289, pp. 95]

[238] Az Orosz Föderáció UNESCO-bizottsága 2020-ban létrehozta a Mesterséges Intelligencia Etikai Bizottságát, amely az UNESCO-n belül tanácsadó testületként szándékozott működni.[292]

[239] A digitális visszaélésre alkalmas MI-technológiákat már 10 éve is exportálták nyugatra. Ld. [293].

[240] 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)

[241] Pl. jó mimikájú humanoid https://promo-bot.ai/robots/robo-c/ (Letöltve 2023. 03. 17)

ellenzéki képviselő mellett tüntetők azonosításától a hadkötelességet elkerülők lefüleléséig.[294] A modell tehát működik, az orosz főváros és Szentpétervár a kínaiénál jóval kisebb IoT kamerás lefedettsége mellett is.[242] Sőt, felzárkózóban van olyan újabb alkalmazások által, mint a pár hónapja bemutatott Oculus[243], 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 döntéstámogatás.[296] 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, sem pedig az akadémiai kutatások nem képesek szerintük garantálni (ezért olyan kevés a szabadalmak száma).[244] A szkepszis ellenére azonban sikeresen alkalmazzák az MI-t, például a KUB-BLA felderítő harci drónokban.[298, pp. 3] Tehát az orosz MI egyelőre nem rikító, 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 [299], és már 2019-től folynak erre irányuló, hivatalosan elismert tesztek.[245] Ez a törekvés éveken át csak erőlködés volt: például, ha valamit tiltottak, akkor más sem működött.[246] 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 leválasztására 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

[242] Moszkvában 213 ezer körül van a számuk, Szentpéterváron ezer főre 12,7 felderítő eszköz jut. Ld. [295].

[243] 2023 februárjában, ld. https://www.interfax.ru/russia/885877, Letöltve: 2023. 03.12.

[244] Az erős korrupciós lehetőségek miatt. Ld. [297].

[245] Ekkor még a felhasználók (állítólag) nem észlelték.[300]

[246] Pl. a Twitter tiltásának egyik próbája az egész internetet belassította. Ld. [301].

vasfüggöny[247] 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 internetleállás becsült kára egy nap alatt 442 millió dollár volt [303, pp. 169], más forrás szerint összesen 861 millió dollárt vesztett az orosz állam a leállítások miatt 2022 első három hónapjában.[304][248] Annyi biztos, hogy hivatalos nyilatkozatok szerint nincs meg az ország internet nélkül.[249] Tehát valószínűsíthető, hogy kerülik 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 a megszerzett adatok elemzése sajnos a polgárokról való adatgyűjtésre is alkalmas, illetve újabb és újabb tartalmak, tartalomtípusok szűréséhez (1-es irány) vezetett. Az ilyen tárgyú törvények egyre több olyan weboldal és szolgáltatás ellen tesznek lehetővé jogi fellépést, melyek a hivatalos állásponttól eltérnek.[250] 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 VPN[251]-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.[252]

A visszaélések mellett szükséges rámutatni, hogy mindebben egy digitális szuverenitási igény csap össze a különféle érdekek térségre 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 [307] vagy a szoftverek függetlenítése[253] irányuló törekvések példái mutatják.

A külső nyomásgyakorlás hátulütője, hogy a belső visszaéléseknél kihasználható: megideologizálhatóak a törvények, ha „ezek ellen a külföldi ügynökök ellen" szólnak. A politikainál

[247] Más neveken „szilíciumfüggöny", „internetes vasfüggöny".[302]

[248] 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.

[249] Lavrov külügyminisztert idézi: [305].

[250] Számtalan példát hoz különböző visszaélésekre a Freedom House civil szervezet jelentése.[306]

[251] Virtual Private Network – ehhez csatlakozva elrejthetők azok a digitális adatok, melyekkel egy szolgáltatás használója beazonosítható.

[252] Federal Government Information System: állami ellenőrzés alá vonja a VPN-identitást és az adatforgalmat is.

[253] Még az MS Windowst is cserélik.[308]

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 [302], a Microsoft pedig nem enged hozzáférést a szervereihez.[309]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 [310], 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.[311]

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 szoftverkalózkodást, a fent említett Microsoft-tiltás miatt. Ehhez adódhat idővel a digitális valuták legalizálódása a nem hivatalos hibrid műveletek támogatásához, és így 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.[254] Ráadásul a cégek sokáig kijátszották, illetve bojkottálták ezeket a rendelkezéseket [303, pp. 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 erőtlen volt 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.[255] 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

[254] 2000-től kezdődött a korlátozások törvényi megalapozása. Ld. [290].

[255] Mivel központi kiberparancsnokság sincs. Ld. [293, pp. 30].

(az utóbbi 10 évben) sokkal több társadalmi feszültséggel jár az a folyamat, amelyben a hatalom el kívánja venni az emberektől, amit használtak és szerettek, és 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 fejlesztésekben. Ez a *gazdasági függőség* (3) magával hozta, hogy a működésbiztonság érdekében érdemesebb 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* [256] (4) is. Érdemes lenne ezeket a tényezőket saját régiónk szempontjából is megvizsgálni.

VI.3.4. Következtetések és az ukrán-orosz konfliktus

Az előző rész 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 részeiként értelmezendőek. 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 több szempontú vizsgálatának módszerét, melyben a személyiségjogi szempontok együtt kezelendőek az állam és a társadalom érdekeivel. (A vállalati szféra szempontjainak kibontására nem volt mód, csupán felvetettem a puha és közepes műveletek nem állami vektorait.) Ezzel a megközelítéssel lehet esély arra, hogy a digitális szuverenitás elérhessen egy elégséges nívót az egyéni, a csoport, a vállalati, a társadalmi és az 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-

[256] [303, pp. 169] Megjegyzendő, hogy erre a kínai beszállítókra áttérés sem megoldás, csak más illesztési problémákat vet fel.

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 küzdőtérré.

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 szembesülnie 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 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 nagyszámú 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.4. A KATONAI INFORMATIKA ÉS AZ INFORMATIKAOKTATÁS LEHETÉSGES VÁLTOZÁSAI

Az eddig felvetett témákat rengeteg területen lehetne még vizsgálni, ezek közül én két olyan témát választottam továbbgondolásra, mely közvetlenül érinti munkámat: a katonai informatikát és a védelmi oktatást.

VI.4.1. A katonai informatika felértékelődése

Alább, részben az informatika fogalombővülését (vagy új fogalom megjelenését) taglaló gondolatmenethez térek vissza (V.3.2.), részben pedig az adaptált katonai informatikához (VI.1.4.). Ez utóbbinál bemutattam, 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 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, (a kibertámadások végrehajtása nem az ő feladatuk), de sokak számára tűnhet úgy, hogy az ilyen harcokat a hátországból vagy védett helyről, „fotelből" lehet elvégezni. Szakmailag hiába evidens, hogy sérülékenységi pontokat eredményez, ha távolról menedzseljük vezetési számítástechnikai rendszereinket, rendszerelemek pedig nem csupán az irányító pontokon lehetnek, tehát a feladat valójában a hadszíntérre rendeli az informatikus katonákat. A támogató állományról jobbára más kép él, és az az attitűd, mely szerint az informatika távol marad a tűzvonaltól, a szakállomány presztízsét szinte teljesen elvette.[257] Azonban a hibrid műveletek fejlődése és a kibertér hadszíntérré válása megindított egy ellentétes irányú folyamatot, melyet az MI bevonása a műveletek különböző szegmenseibe, főleg az irányításba, várhatóan további erősít, mint minden kibertéri hatást (ld. VI.2.). Ez a tendencia várhatóan olyan új elvárásokat fog támasztani a katonai informatika irányába, mely által annak szerepe felértékelődhet.

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. Távolabbra tekintve ugyanis az MI-nyárról említett meglátásom (IV.5.3.) 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 békefelhasználásokként kerülhetnek felhasználásra, pl. hazánkban a KGIR[258]-rendszer ilyen irányú automatizálása képzelhető el, melyhez az SAP cég biztosíthatja az MI-modulokat. Harci alkalmazásokhoz direkt katonai MI-ket 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 szemé-

[257] Sajnos van igazsága ennek a benyomásnak, hiszen egyelőre nem terjedtek el az MH-nál a harctéri IoT, moduláris (pl. Rasbery-alapú) „okosító hardverek" vagy fejlett hálózati technológiák, olyanok, melyek eszközei fizikai jelenlétet igényelnének.

[258] Teljes nevén: Honvédelmi Minisztérium, Költségvetés Gazdálkodási Információs Rendszer.

lyiségi jogok védelme, hanem a hadijog irányelvei adnák (sok más különbség is lesz). Valószínűsíthető, hogy az ilyen rendszerek fejlesztését a világ sok helyén nem katonák, hanem bevizsgált magánvállalkozások vagy állami cégek végzik.

Ám harci helyzetben az ilyen támogató képességek kihasználását katonákra kell bízni. Ehhez pedig kifejezetten erre is kiképzett tisztekre, főtisztekre lesz szükség. Ez, úgy vélem, hogy jóval nagyobb vertikális tanulási lépést igényel (ld. IV.5.), mint amelyet a jelenlegi, hagyományos számítástechnikára szocializálódott vezetői generáció könnyen megléphet. Vagyis a meglévő tapasztalt, képzett és kreatív vezetőket sokkal nehezebb lenne továbbképezni az adatbázisok, a Big Data adatbányászat és az MI-képességek hatékony kihasználására, mint egy eleve számítástechnikára képzett szakállományt. A gépesítés egyfajta kiszámíthatóságot is adhat a műveleti tervezésnek, ami taktikailag igen hátrányos. Ezért szükséges lesz úgy megfogalmazni a tervezési feladatokat, hogy abban a gép generatív moduljai is kihasználásra kerüljenek, ami műszakias gondolkodást igényel. Ehhez járul a különböző harci robotok felprogramozásának tervezése és felügyelete, a harctéri IoT hálózatok adataira elemző szkriptek előállítása, vagy ezek MI-n keresztüli analízisének beállítása, és hasonló újgenerációs harci tevékenységek. Mindezek arra mutatnak, hogy ez a feladatkör úgy lesz erősen katonai jellegű, hogy ugyanakkor magasan képzett informatikusokat igényel.

Vagyis a katonai informatikával kapcsolatban egy paradigmaváltásra lesz szükség. Az új informatikai tisztek valódi vezetéstámogató feladatokat is el kell, hogy lássanak. Így egy új informatikai szakállomány feladata lenne a művelet vezetőinek szándékát a gép számára precízen értelmezhető formába konvertálni, és a kapott választ értelmezni. A jövő az olyan rendszerekben van, melyre jó példa a JADC2[259], vagyis a Közös (összekapcsolt) Összhaderőnemi Vezetés és Irányítási rendszer tervezete. Ez az Egyesült Államok hadereje számára kíván MI által támogatott elemző és tanácsadó funkciókat is szolgáltatni.[312]

Ilyen képesség jó kihasználása azonban nem képzelhető el a mai szokásrendszerben. Az újdonságok a vezetői megközelítés paradigmaváltását is igénylik. Elsősorban szükség lesz egy interakcióra, vagyis a jövőbeni parancsnokoknak engedniük kell például a visszakérdezéseket az informatikai szakemberek részéről. Hiszen azok csak így lesznek képesek a feladatot félreértések idővesztesége 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. Ilyen paradigmaváltás nem igényel nagy vertikális tanulási lépést, ezért ennek elsajátítását reálisnak tartom a máshoz szokott katonák számára is.

[259] Joint All Domain Command and Control – vagyis egy kiterjesztett C2 rendszer.

VI.4.2. Az MI és a védelmi oktatás

Az előző szakaszok gondolatmenete alapján vizsgálni szükséges minden harcoló és támogató állomány MI-vel kapcsolatos kiképzését is, különösen az informatikai (vagy hasonló irányultságú) képzés MI-ismereteinek magasabb szintre emelését. Alább a katonai felsőfokú képzésben megvalósítandó oktatást vizsgálom, de nem azzal a céllal, hogy erre terveket dolgozzak ki, csupán tömören áttekintem ennek lehetséges irányelveit, melyeket mindössze vitaindítónak szánok.

Minden képzésben szükséges lenne már most az MIKT részeit és működését elmagyarázni, hangsúlyt fektetve a félreértések tisztázásának fontosságára (pl. pszeudo-MI, V.1.4.). Az ilyen ismereteket eltérő mélységben kell oktatni a képzések szintje és a szakok érintettsége szerint. Ez az informatikaoktatók feladata. Bizonyos mértékben ez már most is megvalósul, hiszen – óraszámhoz igazítva – jómagam is minden képzésbe igyekszem ezt beépíteni. Hosszabb távon ezt tudatosan bővíteni szükséges, hiszen a hallgatók a most még újnak számító alapismeretekkel tisztában lesznek, ezért több idő marad az általuk nem ismert, de fontos részletekre, vagy kezelésük gyakorlására. A különböző, minden fegyvernem sajátos igényeit kielégítő adaptív és direkt katonai MI-k fejlesztéséhez szükség lesz védelmi vagy katonai tudást is hozzáadni, ehhez pedig az MI és Big Data alapokkal tisztában lévő tisztek kellenek majd minden szakterületen, akik az informatikai fejlesztések számára saját területüket segítenek modellezni, vagy a modellt (programot) ésszerűsíteni, javítani. A legfontosabb persze a megfelelő szemlélet átadása, amellyel utána önképzés által is tudnak fejlődni a végzett kollégák. Ezzel vissza tudok csatolni arra a következtetésre, amely a vázolt MI-nyár kiaknázhatóságára a lakosság MI-használói képzését jelölte meg (IV.5.3.): a védelmi oktatásban ezt kiemelten lenne szükséges végrehajtani.

Külön ki kell térni a nyelvi és egyéb (kép, videó) szolgáltatások használatára, melyeket a fiatal hallgatók saját célra ugyan előszeretettel alkalmaznak, azonban ezek szakszerű promptolását nem ismerik. Ennek oktatását azonban NEM az informatikai vagy infokommunikációs oktatók feladataként képzelem el, hanem mindenhol megjelenő kompetenciaként. Az NPT-promptolás beépítése minden oktatásba okot adna az oktatóállomány számára is a szükséges önképzésre (melyet utána kutatói tevékenységükben is jól használhatnának). Emellett alapot biztosítana a jövőre nézve a folyamatosan fejlődő NPT-rendszerek tesztelésére is. Hiszen annak folyamatosan változó válaszai tendenciózussá is válhatnak, és egy ilyen irányú folyamat védelmi szempontból fontos, ha nagy arányban pl. az országot hátrányosan bemutató válaszokat kezd el adni (akár torzítási hiba miatt, akár információs művelet részeként). Ezt leginkább a katonai kutató állomány jelezheti az illetékesek felé, ha használja az MI-szolgáltatásokat. Mivel

a kutatók számára a vertikális tanulás szinte egy életen át tartó elvárás, ezért ők képesek is ezt elsajátítani és oktatni – ha lesz erre elegendő motivációjuk. Igaz, egyelőre ezek a szolgáltatások meglehetősen sok kompromisszumot igényelnek, és energiát szükséges megismerésükbe fektetni. Jelen fejezet során elemzett változások azonban rámutatnak, hogy minden szakirányon idővel szükséges is lesz beépíteni a számonkérési- és elvárásrendszerbe a digitális források és a nyelvi modellek kritikus és helyes használatát. Ez az információs megtévesztések elleni védettség szempontjából is fontos minden hallgató számára, de kritikus jelentőségű a védelmi szférában tanulóknak, hogy pl. rutinosan tudjanak megkülönböztetni valós és álhíreket. Nem vagyok optimista azt illetően, hogy ezek a közeljövőben beépülnek minden tanrendbe. De minél előbb hajlunk az oktatás ilyen irányú alakítása felé, annál jobb pozícióba kerül a védelmi oktatás és az ország.

Végül említeni kell a harci MI-ágensek használatának oktatását. Ezekre különleges kiképzések során már jelenleg is felkészítik az erre kijelölt állományt. 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.[260] Egyelőre túl távolinak tűnik magyar vezetéstámogató MI, illetve fejlett robotok megjelenése csapatainknál, de a téma folyamatos evidenciában tartásához beadandó feladatok és szakdolgozatok formájában lehetősége van az oktatóknak, amit karunkon végre is hajtanak.

VI.5. RÉSZÖSSZEFOGLALÁS: VÉDELMI ASPEKTUSOK

Összegzés. A fejezet vizsgálataival sikerült a korábbi kutatások gondolatait befejezve számos szempontból feltérképezni az MI védelmi aspektusait. Ehhez, a K6 kérdésre válaszolva a bevezetőben kijelöltem azt a négy témát, melyeket körbejárva a C2 cél megvalósult. Megfogalmaztam azt a négy irányt is, melyek mentén a H2 hipotézis bizonyítását végre akartam hajtani. A feltevés bizonyítását majd az összegzésben írom le, összegzésként elegendő annyi, hogy az ezekhez szükséges elemzéseket sikerült integrálni az alfejezetekbe. Ezek közül az első irány inkább hadtudományi aspektusból közelített, és így mutatta be az erőérvényesítés megváltozását, valamint azt, hogy ezt hogyan katalizálja az MI. A második alfejezet a korábbi technikai ismertetésekre támaszkodva adott rövid körképet az MI kibertéri használatáról. A harmadik részkutatás a technológiákkal való visszaélés felől közelítve mutatott rá a puha műveletek jelentőségére. Az utolsó alfejezet a hipotézis két következményét elemezte röviden.

[260] Megjegyzés: a NKE HHK Logisztika Tanszékével közös újgenerációs adatelemzési kurzust tervezünk.

P2-vel kapcsolatos következtetések

(P1 vizsgálata az előző fejezetben lezárult)

R6.1. Az erőérvényesítésben sokrétű, komplex paradigmaváltás folyik (VI.1.1.).

R6.2. Az R6.1. kezelésére új modellek és terminológiák is kellenek, jelen kutatásból az alábbiak említendők:

R6.2-A A virtuális erőközpontok modellje rávilágít arra, hogy az erőérvényesítésben megjelenő új paradigmákhoz jól alkalmazható az MI, sőt generálja is az új irányokba mutató tendenciákat (VI.1.1-3.).

R6.2-B A direkt és adaptált védelmi informatika fogalmi szétválasztása (VI.1.4.) segít a fejlesztések katonai sajátosságait besorolhatóbbá tenni.

R6.2-C A kompozit (hibrid) hidegháború jellemzőinek (VI.1.5.) segítségével jobban megragadható a mai helyzet, akár az orosz–ukrán háború esetében (VI.3.3-4.), akár egyéb konfliktusokban.

R6.3. Az autonóm fegyvereket (az MI kemény műveleti alkalmazását) sokan ellenzik (VI.1.5., VI.2.4.).

R6.4. Az R6.3 folyományaként a puha műveletek előtérbe kerülését a társadalmi elvárások jobban pártolják, mint az erőszakot (VI.1.1.).

R6.5. Az MI megsokszorozza a kibertéri lehetőségeket és veszélyeket (VI.2.1-2).

R6.6. Az affektív számítástechnika eredményei a védelmi szféra számos területén jól, a puha műveletekben kiemelkedően alkalmazhatóak (VI.2.3.).

R6.7. Jelenleg és a közeljövőben az MI a puha műveletek területén jóval veszélyesebb, mint az autonóm fegyverrendszerek robotjaiban (ezeket a távirányítású robotokkal vagy robbanó-, vegyi- vagy biológiai fegyverekkel összemérve) (VI.2.4.).

R6.7-A: Az elszabadult technológiák veszélye az MI nélkül is fennáll, és a vékony MI-kben épp úgy javítható a hiba, mint a hagyományosan kódolt automatikákban.

R6.7-B: A hagyományos rendszerek sokkal veszélyesebbek részleges meghibásodás esetén, mivel ilyenkor részlegesen működnek, viszont egy feltanított neuronháló elromlása kezelhetőbb.

R6.8: A katonai és védelmi informatika korábbi szolgáltató-üzemeltető, vagyis támogatói szerepe részben a kibertér, de sokkal inkább majd az MI hatására felértékelődik (VI.4.1.).

ÖSSZEGZÉS, EREDMÉNYEK, HASZNOSÍTHATÓSÁG

A záró alfejezetek előtt szeretnék köszönetet mondani mindazoknak, akik bármilyen formában hozzájárultak dolgozatom létrejöttéhez.

ÖSSZESÍTETT KÖVETKEZTETÉSEK

A hipotézisek bizonyításához elsősorban a fejezetek végén található részkövetkeztetéseket használom fel. A fontosabb gondolatokhoz megjelölöm a vonatkozó fejezetszámokat is. Az igazolások logikája úgy véltem, követhetőbb, ha vizuális ábrázolásokkal segítem az összefüggések egymáshoz való viszonyának áttekintését. Ezért gondolattérképen is ábrázoltam a következtetéssorok lényegét,[261] melyeket a magyarázatban helytakarékosan (19. és 20. ábra), a mellékletek között kinagyítva ismertetek (21. és 22. ábra), ahol a forrásfejezetekre az eltérő színek utalnak.

Az E1 eredményhez vezető következtetés-sorozat

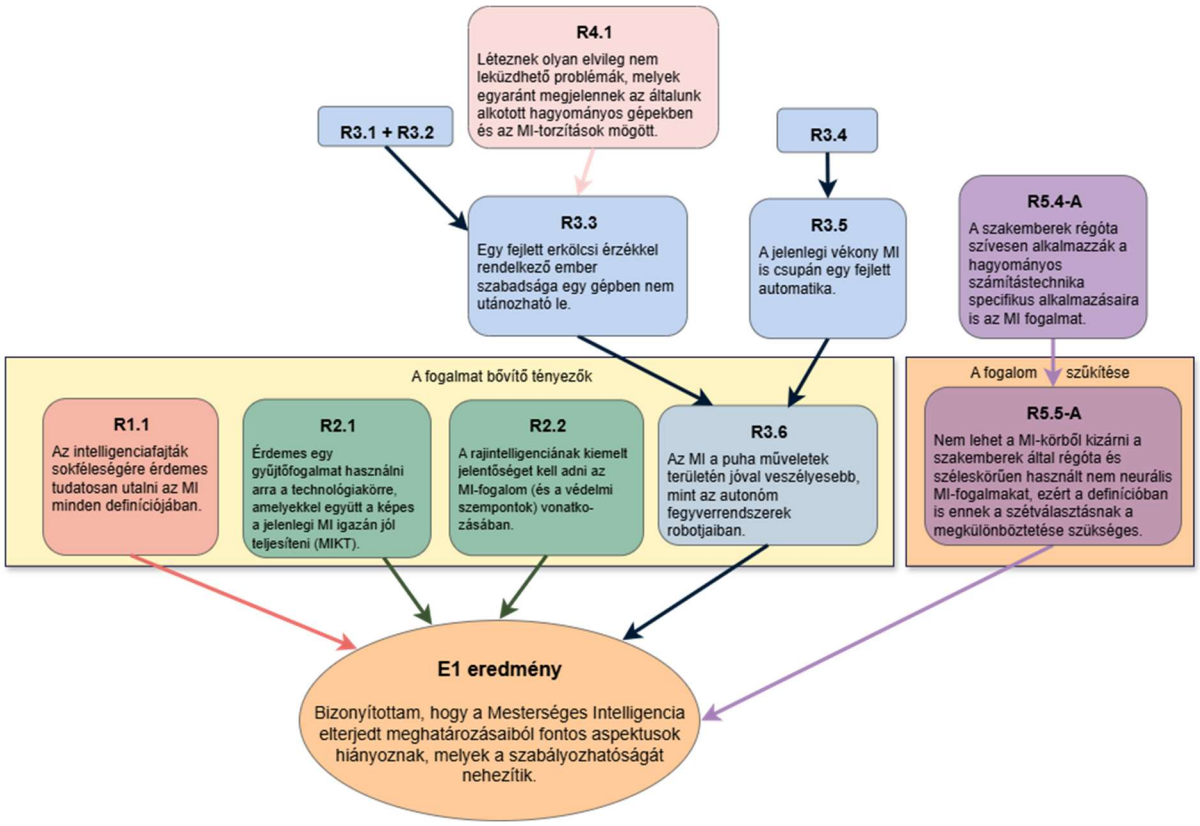
Bizonyítandó volt, hogy *az MI jelenlegi meghatározásaiból fontos aspektusok hiányoznak* (H1). Ennek igazolásához C1 célt arra irányoztam, hogy az MI-fogalom túl szűk és túl tág használatait elemezzem. Mind a két problémára tártam fel olyan példákat, melyeket orvosolni szükséges. A javaslataimban konkrét kifejezéseket ajánlok, néha alternatív megoldásokat is említek, ám ezek mindig inkább egy adott gondolatkörre utalnak. Vagyis nem ezen kifejezések használatát javaslom, hanem bizonyos jellemzők vagy elvek megjelenítését látom szükségesnek egy jobb MI-definíció érdekében. (Javasoltam bizonyos kifejezések kerülését is a meghatározásban, de ezek áttekintését nem célozta a kutatás, ezért itt ezeket nem tárgyalom.) A megjelenítendő szavakat és azok indoklását a 3. és 4. táblázatban foglaltam össze az V.4.2. fejezetben. Ezek az összefoglalások új MI-fogalmak megszövegezésére irányultak, alább viszont a H1 bizonyításaként fogalmazom meg az érvek és gondolatok összegzését. A kutatásban öt olyan kifejezést (tényezőt) sikerült kimutatnom, melyek nélkül az MI-definíciók félreérthetőek, tehát igazolják a feltételezést (ld. 19. ábra). Ebből négy az MI-fogalomból *hiányzó* aspektus:

- **A sok fajta intelligencia** (R.1.1.) említésének hiányát tartom a legsúlyosabb fogalmi hibának. Ez ugyanis azt a régi filozófiai-antropológiai tévedést erősíti, miszerint az intelligencia az okossággal azonosítható. Ezt a korai hibát a tudomány már kiköszörülte,

[261] Ezekben nem írom ki részletesen az idézett részkövetkeztetéseket, csak a lényegüket; sőt az áttekinthetőség kedvéért némelyiknek csupán a számát adom meg az ábrákon.

számos más intelligenciatípus utánzásának fejlesztése zajlik, ezért a kifejezéskör elhagyása technológiailag is megtévesztő (és diszkrimináló) (I.2. , I.3.).

- **A technológiakör az MI mögött** (R2.1.) azért lenne belefogalmazandó egy világos definícióba, mivel nagyon lényeges használhatóságbeli eltérés van egy mobiltelefon arcfelismerő MI-ágense között és olyan MI között, mely erős felhőszervereken fut, óriási adattóból meríti tudását, melyet IoT eszközök milliói is táplálnak. Ez utóbbi technológiakörre én az MIKT[262] betűszót javasoltam (II.1.).



19. ábra: A H1 bizonyításának gondolattérképe (saját készítés – nagyítva ld. 21. ábra)

- **A biológiai együttműködések utánzására valamilyen utalás** (R2.2.) azért lenne fontos, hogy a fogalom jól fedje le a lényegesen eltérő fejlesztési irányokat. Márpedig az élő szervezetekben végbemenő folyamatok, vagy bizonyos fajok egyedei között kialakult együttműködések modellezése lényeges eltérést mutat a centrális MI (pl. MIKT) működésétől, hiszen pl. elosztott (nem-centrális) intelligens rendszereket is lehetővé tesz. Én a rajintelligencia említését javasoltam a meghatározásba, annak ismertebb volta, elterjedtsége, valamint a szó tömörsége miatt (ld. II.3.3. – a sokkal ideillőbb bionika kifejezést ismeretlensége miatti félreérthetősége okán vetettem el).

[262] Mesterséges Intelligencia és Kapcsolódó Technológiák

- **Automatika** (R3.6.). Használatát inkább az autonómia kifejezés kerülése miatt tartom fontosnak, az MI-től való félelem csökkentése érdekében. Ugyanis R3.5 alapján a jelenlegi vékony MI-rendszerek még hipotetikusan[263] is csupán az automatika egy magasabb fokát képesek megvalósítani. Ez utóbbi állítás a nemzetközi J3016 szabvány (III.3.2., III.4.1.) mellett, R3.4.-re épül (ld. alább), elvi alapja azonban az R3.3. Ez utóbbi szerint ugyanis az ilyen gépek elvileg sem képesek az emberhez hasonló autonóm döntésre. Ezt az elvi lehetetlenséget részben R3.1-re alapoztam, amelynél elemzésekkel támasztottam alá, hogy alapvető különbségek azonosíthatóak az emberi és gépi autonómia között[264] (III.5.). De R3.3.-at támasztja alá a később kimondott R4.1. is, hiszen ez alapján léteznek *elvileg* leküzdhetetlen problémák (IV.2-3.), más szóval az ember gépbe másolása nem csupán „gyakorlatilag nem sikerült" ezekkel a rendszerekkel. Visszatérve az említett R3.4.-re, ez a megállapítás ugyanerre a különbségre másképp mutat rá,[265] amikor azt mondta ki, hogy a gépeknél az autonómiaszinteket *a beépített MI szintjéhez* kellene kapcsolni[266] (III.4.1.), ellenben az emberek esetében az autonómia szintjei *etikai* alapon fejthetőek vissza. (III.1-2.)

Az ötödik megjelenítendő tényező nem hiányosságot azonosít, hanem arra fókuszál, amit *kiküszöbölni* kellene az MI-fogalomból. Ez a korrekció az MI-kifejezés ellentmondásos (zavaros) használatát kívánja helyretenni az alábbi ajánlással:

- **A neurális MI** (R5.5-A) javaslata valamilyen előtaggal javasolja megkülönböztetni a neurális hálózaton alapuló, feketedoboz jelleggel működő MI-ket azoktól a hagyományos rendszerektől, melyeket különböző okokból is MI-nek hívnak, hisznek, vagy annak állítanak be (holott pszeudo-tanulással működnek vagy tanulni sem képesek, ld. V.1.). A hagyományos kódokat is lehet előtaggal jelölni és „determinisztikus MI-ágens"-nek vagy „nem neurális MI"-nek hívni. Mindezt elsősorban szabályozások megfogalmazásaiban lenne fontos használni.

Ezeken kívül további négy kifejezést javasoltam megemlíteni a precízebb megfogalmazás érdekében, amikor van erre lehetőség (R2.4., R4.3. és R5.6. alapján). Ezek: a *szinergia*, *kogní-*

[263] Ezzel a még forgalomba nem került „ötödik szintű önvezető autók" lehetőségére utalok.

[264] Ez az állítás az embergép koncepció elterjedtsége miatt nem evidencia.

[265] Az, hogy a gépek valójában nem képesek emberibb autonómiára, azt a III. fejezet egésze bizonyítja. Nagyságrendekkel komplexebbnek kellene lenniük az ezt célzó gépeknek, hogy ilyesmi egyáltalán felmerülhessen, mivel az ember náluk nagyságrendekkel komplexebb.

[266] Annak érdekében, hogy a különféle rendszerek biztonsági szempontból jól besorolhatóak legyenek.

ció, *leképezés*, *számítástechnika* (az informatika helyett!). Mivel ezek nem kifejezetten hiányoznak a kifejezésből, csupán pontosítják, ezért nem tartoznak a bizonyításhoz, tehát magyarázatuk itt mellőzhető. (Használatuk indoklása és az alfejezetek, ahol ezek hasznosságát tárgyaltam, megtalálható a 3. és 4. táblázatban, V.4.2.-ben.)

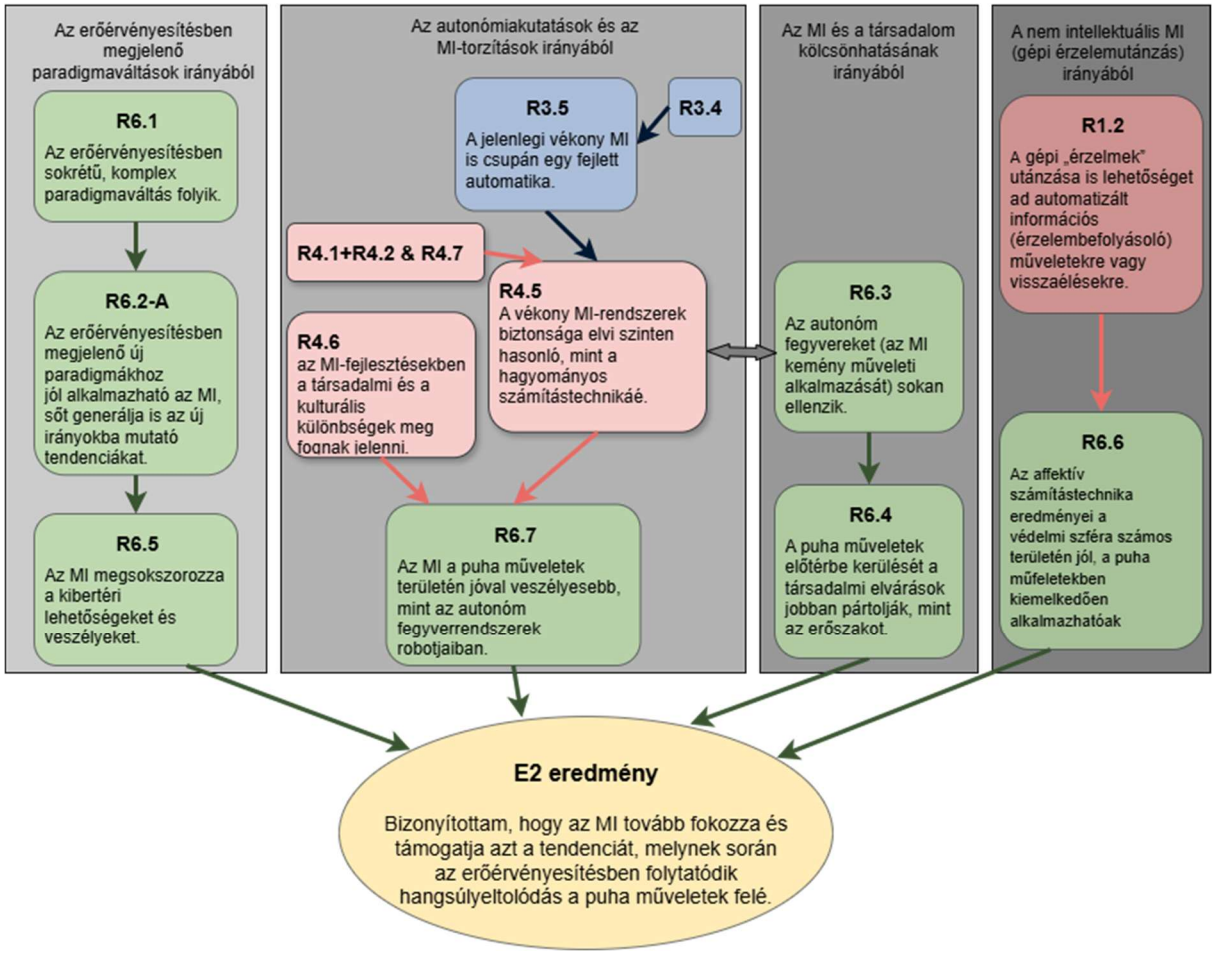
Az E2 eredményhez vezető következtetés-sorozat

Bizonyítandó volt, hogy *az MI tovább fokozza és támogatja azt a tendenciát, melynek során az erőérvényesítésben folytatódik hangsúlyeltolódás a puha műveletek felé* (H2). Ennek igazolásához C2 célt arra irányoztam, hogy az erőérvényesítés új tendenciáit és azok MI-vel való kapcsolatát elemezve meghatározzam az MI védelmi szempontból fontos aspektusait. A H2 bizonyítását négy irányból hajtottam végre: (1) Az erőérvényesítésben megjelenő paradigmaváltások irányából, (2) Az autonómiakutatások és MI-torzítások irányából, (3) Az MI és a társadalom kölcsönhatásának irányából, (4) A nem intellektuális MI (gépi érzelemutánzás) irányából (ld. 20. ábra).

(1) Az első gondolati szál érvei háromféle konklúzióra alapulnak. Az R6.1. kimondja, hogy egy paradigmaváltás valójában az MI nélkül is már zajlik egy ideje az erőérvényesítésben, és be is mutattam ennek főbb vonásait (VI.1.1.). Viszont R6.2-A alapján erre a tendenciára igen jól illeszkedik az MI képességrendszere; nagyon jól alkalmazható ilyen frontokon, sőt megjelenésével az MI tovább katalizálja a megindult folyamatokat (VI.1.1-3.). Az előzőekből is szinte következik, de külön elemeztem is (VI.2.1-2), hogy hogyan sokszorozza meg a kibertér lehetőségeit az MI – annak veszélyeit is jelentősen növelve.

(2) A második érvrendszer összetett. Több, egymásra épülő részkutatás alapján állt elő az az érv, mely R6.7.-ban azt mondja ki, hogy jelenleg és a közeljövőben a puha műveletek területén jóval veszélyesebb az MI, mint az autonóm fegyverrendszerek robotjaiban (ezeket a távirányítású robotokkal vagy robbanó-, vegyi- vagy biológiai fegyverekkel összemérve) (VI.2.4.). Két gondolatmenet irányából jutottam erre a konklúzióra. Az egyik egy társadalmi oldalról jövő meglátás (R4.6.), melyben arra mutatok rá, hogy az MI-fejlesztésekben a kulturális és társadalmi különbségek meg fognak jelenni (IV.2.2.(3)). Ezek az eltérések eltérő etikájú kibertámadásokban is megnyilvánulhatnak, aminek veszélyei szinte prognosztizálhatatlanok. A másik gondolati szálon (R4.5.) azt mutattam ki, hogy a vékony MI-rendszerek biztonsága *elvi* szinten hasonló, mint a hagyományos számítástechnikáé (gyakorlati szinten

nagyon nem hasonlóak, IV.2-3.). Erre az álláspontra egyrészt az autonómiavizsgálatok alapján jutottam, melyek során kiderült, hogy a jelenlegi vékony MI is csupán egy fejlett automatika (R3.5.[267]), másrészt ezt megerősítettek az MI-torzítások vizsgálatai[268] is (IV.2-3).



XX. ábra: A H2 bizonyításának gondolattérképe (saját készítés – nagyítva ld. IV. ábra)

(3) A harmadik irány felől úgy mutatok rá a problémára R6.4.-ben, hogy a (főleg nyugati) világ tendenciái az „erőszakmentesség" irányába mutatnak. Ezekhez a tendenciákhoz a puha műveletek illeszkednek legjobban, minden szempontból (VI.1.1.). Az MI részéről ezt erősíti R6.3. szerint az a politikai társadalmi tiltakozás, az a sokkal nagyobb figyelem és félelem, mely az MI keményműveleti bevetését, vagyis az autonóm fegyverrendszereket („gyilkos robotokat") övezi – szemben a puha műveleti MI-használattal, amiről nem szinte is hallani

[267] Az R3.5-höz vezető gondolatok összegzését ld. E1 eredmény lábjegyzetében.

[268] Az R4.5.-re ebből az irányból pl. R4.2. utal, vagyis, hogy az óriásivá növekvő és bonyolódó hagyományos rendszerek és az MI átláthatósága egymás felé konvergál (IV.4.), valamint R4.1. is, melyben azt foglaltam össze, hogy léteznek olyan *elvileg* nem leküzdhető problémák, melyek egyaránt megjelennek az általunk alkotott hagyományos gépekben és az MI-torzítások mögött (IV.2-3.). Ezt másképp ragadja meg R4.7., miszerint egyaránt csak teszteléssel lehet hibát keresni mind a robusztus hagyományos, mind az MI-rendszerekben (IV.2-3.).

(VI.1.5., VI.2.4.). Ez utóbbi a közhiedelmet egyébként a (2)-es irány R4.5.-ös gondolata is cáfolja, hiszen ha mind a régi nagy rendszerek, mind az új neurális rendszerek csak teszteléssel, statisztikai értékkel jellemezve mondhatók biztonságosnak, akkor a gyakorlati nehézségek, melyek egyelőre a hagyományos rendszerek mellett szólnak, csupán a fejlődés jelen éveire jellemzőek, nem pedig *elvileg* „gyilkosok" az MI-robotok (IV.2-3.).

(4) A negyedik irány ezeket az érveket azzal egészíti ki, hogy az affektív számítástechnika eredményei a védelmi szféra számos területén jól, a puha műveletekben viszont kiemelkedően alkalmazhatóak, amint arra R6.6. rámutat (VI.2.3.). Ehhez a kezdeti vizsgálatokra utalok vissza, ahol a technológia alapjai mellet R1.2 hangsúlyozta, hogy nem csupán a gépi „okosság" alkalmazható támadó, megfigyelő, védekező vagy egyéb módon, hanem a gépi „érzelmek" utánzása is lehetőséget ad automatizált információs (érzelembefolyásoló) műveletekre. (I.4.)

TUDOMÁNYOS EREDMÉNYEK

A hipotézisek bizonyításával a kutatás az alábbi tudományos eredményeket érte el:

E1 eredmény:

Bizonyítottam, hogy a Mesterséges Intelligencia elterjedt meghatározásaiból fontos aspektusok hiányoznak, melyek a szabályozhatóságát nehezítik.

E2 eredmény:

Bizonyítottam, hogy az MI tovább fokozza és támogatja azt a tendenciát, melynek során az erőérvényesítésben folytatódik hangsúlyeltolódás a puha műveletek felé.

JAVASLATOK A KUTATÁS FELHASZNÁLÁSÁRA

A tanulmány sokrétű vizsgálatai, és a több mint ötven részkövetkeztetés számos konklúziót enged megfogalmazni, de remélhetőleg az alábbiak is elegendőek a kutatás gyakorlatias oldalának szemléltetéséhez. Ezeket – a fontosabbnak ítélt – felhasználási módokat négy csoportba rendezve ismertetem (F1-4), azzal a megjegyzéssel, hogy egy-egy felhasználási mód több csoportba is besorolható lenne. Az összeállításhoz a két eredmény ilyen irányú átgondolásán túl a részeredményeket is alapul vettem. Mindezek előtt azonban két olyan javaslatot mutatok be, melyekhez összetettebb következtetések vezettek.

Egy konklúzió és egy sejtés

A két alábbi praktikus következmény közül az első egy olyan tényt akar hangsúlyozni, amely a szakmabelieknek talán evidens, ám a gyakorlati tapasztalat szerint mégsem elfogadott széleskörűen. A másik egy olyan sejtés, melynek bizonyítása még nem teljes, de az összeállított érvrendszer megfelelő kiindulás a további vizsgálatokhoz.

Konklúzió: Alátámasztottam, hogy az MIKT és egyéb paradigmaváltó technológiák oktatásának beépítése elengedhetetlen a katonai és védelmi képzésbe.

A kutatás oktatási vonatkozású meglátásai úgy vélem, hasznos érvekkel járulhatnak hozzá az MI-oktatás hazai prioritásának növeléséhez, valamint a társadalommal és az oktatókkal való elfogadtatásához. (Ezért közlöm annak ellenére, hogy maga a következtetés szakmailag evidens.) Az oktatás területét személyes érintettségem miatt sok helyen megemlítettem: a IV., V. és VI. fejezet vizsgálataiban elméleti szinten elemeztem, közvetlen gyakorlati alkalmazhatósága pedig a II. fejezetnek és az V. fejezet 3. táblázat oktatási oszlopának van.[269] Itt az elméleti vizsgálatok érveit úgy összegzem, hogy négy irányból támassza alá a téma fontosságát.

1. **Az autodidakta tanulás nem működik az MI és egyéb új paradigmák esetében.** Az R4.13. (a vertikális tanulás modellje) kimondja, hogy az ember nehezen vált szemléletet. Márpedig a technológiai paradigmaváltások jó kihasználása (sokszor csupán a használata) nem lehetséges e nélkül, vagyis pusztán többletinformáció elsajátítása által. Természetesen ehhez az új szemléletet átadni képes pedagógusok is kellenek. Viszont tudomásul kell azt is venni, hogy amikor a horizontális tanulás helyett vertikális szükséges, az gyorsan, tanfolyami szinteken nemigen valósítható meg, mivel ez, az ott elérhetőnél sokkal több lelki energiát, inspirációt igényel. Vagyis a hallgatókban az új paradigmák megfelelő szintű értése és alkalmazási képessége (illetve annak csírája) csupán több hónapos, esetenként több éves, szemléletet is oktató képzések során jöhet létre (IV.5.1.).
2. **A tudományok konvergenciája miatt minden diplomás számára szükséges az MI alapszintű ismerete** a saját célrendszerek használatához és alakításához.[270] Az R5.2. szerint minden tudomány konvergál egymás felé az MI-ben (V.3.4.), ezért minden felsőfokú oktatásba beépítendő a neurális hálózatok alapjainak ismertetése és minél többrétű használatuk gyakorlati bemutatása ((V.3.4.), erre az R5.10. is rámutat. Viszont más szempontból támasztja ezt alá R5.9., miszerint az a tény, hogy a technikai fejlesztések olyan szintet

[269] Ez az első tananyag alapjaként szolgál, a második tematikákhoz nyújthat alapot.

[270] Vagy akár saját tudománya felől az MI-modellekhez való hozzájáruláshoz.

értek el, ahol be kell várniuk a humán- és egyéb (pl. védelmi) tudományokban kidolgozandó modellek megjelenését lényegi továbblépéseikhez (V.3.1.). Tehát az MI biztonságos terjedéséhez és fejlődéséhez minden tudomány képviselőjére szükség van.

3. **Ha nem lesz rövidesen technikai áttörés, akkor a stagnálást kell kihasználni, ha pedig lesz, akkor arra fel kell készíteni a polgárokat.** Az R4.14. fogalmaz úgy, hogy „az MI hosszú nyara várható", vagyis terjedni fog az MI és jelentős bevételeket termel, de fejlődésében a felhasználókat is érintő, diszruptív változásra véleményem szerint nem igazán lehet egy ideig számítani (IV.4.). Ez a helyzet R4.15. alapján kiaknázható, ha a lakosságot tudatosan inspirálják a jelenlegi MI használatának, paradigmáinak elsajátítására. (Ez akkor sem felesleges, ha mégis egy új MI- paradigma köszönt be, mivel szinte bizonyos, hogy az sok mindenben örökli a jelenlegi technológia számos jellemzőjét.) (IV.5.3.)
4. **A katonai és a védelmi szférában más okokból is nagy jelentőségű az MI-alapok általános oktatása.** Az E2 eredményben bemutatott érvrendszer gondolatai alapján (a védelmi- és katonai vonalon folyó paradigmaváltás miatt), minden katonának, de különösen az infokommunikációs és informatikus tiszti képzés számára kiemelkedően fontos az MI minél mélyebb szintű megtanítása (VI.4.2.).

Sejtés: Nem vetnek fel alapvetően újszerű etikai problémát a jelenlegi MI-modellek és az irányelvek alapján megvalósuló, sokak által támadott „autonóm fegyverrendszerek".

Ebben a sejtésben a filozófiai etika terén fogalmaztam meg egy meglátást a két eredménynél már felhasznált részkutatások gondolataira alapozva, ehhez másképp rendszereztem bizonyos autonómiával és a gépek hibázásával kapcsolatos részkövetkeztetéseket. A sejtés gyakorlati jelentőségére az R4.10. mutat rá, mely szerint az MI-autonómia biztonságtalanságához kapcsolódó általános vélemény a jövőben még akkor is hátráltatni fogja az MI-be vetett bizalmat, amikor a hagyományos és az MI rendszerek biztonsága között statisztikai szempontból már alig lesz eltérés (IV.3.). A bizonyításhoz kétféle megközelítést használok:

1. Először is az arra irányuló érveket gyűjtöttem össze, melyek az MI képességeiről alkotott túlzó megközelítéseket cáfolják.
    - Az R4.9-A alapján a jelenlegi vékony MI csak az alkotók akarata, vagy sorozatos, nagyarányú gondatlansága miatt lehet képes az emberi szándékkal ellentétesen tevékenykedni, és nem képes „csak úgy" önállósulni, netán tudatra ébredni (IV.3.).
    - A R6.7-B szerint a hagyományos rendszerek sokkal veszélyesebbek részletes meghibásodás esetén, mivel ilyenkor részlegesen működnek, viszont egy feltanított neuronháló

vagy egyszerűen nem működik, ha neuronjainak egy részét fizikailag elveszíti, vagy pedig képes helyrehozni magát, vagy leállítani a veszélyes képességeket.

- Vagyis az „elszabadult technológiák veszélye” az MI nélkül is fennáll R6.7-A alapján, sőt a vékony MI-kben jobban javítható a hiba, mint a hagyományosan kódolt automatikákban (VI.2.4.).

2. Az érvek másik köre arra mutat rá, hogy a hagyományos rendszerek nem annyival megbízhatóbbak, hogy indokolt legyen velük szemben az MI bizonyos fokú „démonizálása”.

- Egymás felé konvergál az R4.2. szerint a hagyományos rendszerek bonyolódása és az MI átláthatósága (IV.4.), vagyis biztonsági szempontból csökken az eltérés közöttük, amint azt R4.5. is megállapítja (bár gyakorlati szempontból egyelőre valóban lényegesen jobbak a hagyományos rendszerek).
  - Ezzel összefüggésben már jelenleg is zajlik a „biztonságok közeledése”, hiszen az R4.7.-ben kimondott gyakorlati problémák lényege, hogy mindkét rendszertípus esetén ugyanúgy, tehát csakis teszteléssekkel oldható meg a biztonság, pontosabban annak egy statisztikailag meghatározott foka (IV.2-3.).
  - A hagyományos rendszerek gyakorlati előnyét relativizálja, hogy az R4.1. szerint léteznek olyan elvileg nem leküzdhető problémák, melyek egyaránt megjelennek a hagyományos gépekben is és az MI-torzítások mögött is (IV.2-3.). Ez is arra mutat rá, hogy a kétféle rendszer alapvetően hasonló gátakkal küzd.

Összegzésként: a hagyományos számítástechnika és az MI nincsenek éles oppozícióban. Az új paradigma (MI) fenti megközelítése által az válik világossá, hogy a gépiesítés új korszaka „csupán” a szabályok újraírását követeli meg, de nem jött létre egy szabályozhatatlan, autonóm, teremtője ellen fellázadó evolúciós lépcső.

A P1 (fogalmi) problémakör elemzéseinek lehetséges felhasználásai

F1.1: Az E1 tétel alapján a tanulmány MI-ben kimutatott hiányossági körei felhasználhatóak

- Szabályozóknál: a megszövegezés segít csökkenteni a kiskapuk számát a szabályozatlan területek csökkentésével.
- A védelmi tervezésben: kezelhetőbbé teszi az MI olyan aspektusait, melyeket a jelenlegi meghatározások nem vesznek figyelembe
- Az oktatásban: segíti az MI lényegének megismertetését, valamint a használatához szükséges szemlélet kialakítását. Ez a katonai képzések számára is kulcsfontosságú. Az oktatást a védelmi rendszer részének tekintve minden területen nagyon fontos.

F1.2: A tanulmány során kimutatott hiányosságok és ellentmondásosságok alapján javaslatot tettem az MI új meghatározására egy bővebb és egy tömörebb verzióban, ezek használatát a forrás feltüntetésével engedélyezem (V.4.). Ez a fogalom néhány évig remélhetőleg használható az F1.1-nél felsorolt területeken.

F1.3: Az MI számos felosztási és rendszerezési lehetőségének összegyűjtése, valamint a besorolási mátrix (V.4.1.) hasznosíthatóak

- szabályozások kidolgozása során;
- oktatásban;
- a fogalom további tökéletesítésekor.

F1.4: Kimutattam, hogy az „informatika" fogalmának tartalma is változhat az MI miatt, sőt lényegesen is átalakulhat (V.3.). Ez a meglátás a védelmi használaton túl az MI-technológiához kapcsolódó egyetemi képzések során lehet fontos, mivel az MI jelenlegi, technika-centrikus oktatását szélesebb elméleti és gyakorlati perspektívákba emelheti.

A P2 (védelmi) problémakör elemzéseinek lehetséges felhasználásai

F2.1: Az E2 eredmény védelmi stratégiai, szabályozási, gazdasági és oktatási területen egyaránt hasznosítható meglátást ad a döntéshozók és a szakemberek kezébe.

F.2: Az E2 eredmény alapján a puha műveleti hangsúly kezeléséhez a hadtudománynak új modellekre és fogalmakra van szüksége. Ezek közül itt kidolgozásra kerültek (VI.):

- A kompozit (hibrid) hidegháború jellemzése a világban zajló folyamatok leírására.
- A virtuális erőközpontok teóriája, valamint a virtuális erőérvényesítés rendszertani modellje, mely a lehetséges védelmi vektorok és sérülékenységi területek azonosításához lehet hasznos.
- Az adaptált és a direkt katonai informatika megkülönböztetése, mely a külső beszerzések és a saját fejlesztések arányait és irányait pontosíthatja.
- A digitális visszaélések eddiginél sokkal átfogóbb rendszerezése, elemzése és gyakorlati példán való bemutatása a számos hadtudományi felhasználási lehetőségen túl a jogvédők számára is hasznos perspektívatágításként szolgálhat.

F2.3: A katonai és védelmi szakinformatika új helyének felvetése, a kidolgozott modellek, és a problémák összefoglalása hosszútávon segíthet a szakemberhiány leküzdésében is (VI.4.).

Oktatási vonatkozású felhasználások

F3.1: Elsősorban az I-II. fejezet került úgy kidolgozásra, hogy abból jegyzet vagy tankönyv készülhessen különféle kurzusok számára, de ennek bővebb verzióiba a tanulmány többi fejezetéből válogatott szemelvények megjelenítését is tervezem.

F3.2: Bevezettem a pszeudo-tanulás, a pszeudo-MI és a pszeudo-autonómia fogalmait (V.1.4.), melyek a népszerűsítő anyagokban és az alapozó oktatásban hasznosíthatóak.

F3.3: Az R4.13 alapján a horizontális és vertikális tanulás szétválasztásának (IV.5.1.) figyelembevétele sokat segíthetne a felnőttképzések, illetve továbbképzések elvárásrendszerének átdolgozásakor, az ilyen képzések tervezésekor (pl., hogy ne tanfolyamokban gondolkodjunk, ha szemléletváltásra van a képzendő személyeknek szüksége).

F3.4: Kimutattam, hogy nem lehet az MI széleskörű társadalmi integrációját a fejlődési sebességéhez mérhető ütemben végrehajtani (R4.15.), ami az oktatástervezésben lehet hasznos (IV.5.2-3.).

Egyéb felhasználások

F4.1: Az autonómiával kapcsolatos sejtés – teljes igazolása esetén – jól hasznosítható az R4.10. szerint az MI-vel kapcsolatos félelmek eloszlatásában és a bizalom növelésében, amely számos módon járulhatna hozzá a világ problémáinak megoldásához.

F4.2: Az emberi és főleg a gépi autonómia új típusú, használhatóbb felosztásai (III.) használhatóak az MI-rendszerek osztályozásához és szabályzásához (véleményem szerint jobban, mint a jelenlegiek), emellett az információbiztonsági kérdésekhez is – akár olyan konkrét esetekben, mint pl., hogy a magas rendelkezésre állású rendszerek (HAS) legfeljebb delta szintű, azaz magára hagyható komplex autonómiával valósíthatóak meg („gépi user error” nélkül) (R3.9., III.4.3.).

F4.3: Az információbiztonság vonatkozásában az a megállapítás is hasznosítható, hogy az óriásivá növekvő és bonyolódó hagyományos rendszerek és az MI átláthatósága egymás felé konvergál (R4.2., IV.4.), hiszen ezzel a rendszerek biztonsági megközelítéseiben is elképzelhető egy szemléletváltás szükségessége (R4.8.). Emiatt minden definícióban és szabályzásban a két rendszer együtt kezelése és ugyanakkor eliminálása egyszerre szükséges.

F4.4: Kimutatva, hogy minden tudomány konvergál egymás felé az MI-ben (R5.2., V.3.4.), rendszerszinten vált láthatóvá, hogy az MI-alapok ismerete minden tudományterületen fontos (akár kutatói, akár alkalmazói szinten). Ez jól felhasználható a humánerőforrás-

menedzsment elvárásaiban mind állami, mind pedig a magánszférában, valamint az oktatásszervezésben, hiszen R5.10. alapján a felsőoktatásban mindenhol szükséges a képzésekbe beépíteni az MI projektekben való részvétel megalapozását (V.3.4.).

F4.5: A terminológiai degradációs modell (R5.3., V.1-2.) hozzá kíván járulni a marketing etikusabbá tételéhez, illetve a vásárlói tudatosítás elősegítéséhez.

F4.7: A gazdasági szféra számára lehet jól hasznosítható modell a fejlődés több dimenziós (horizontális és vertikális) leírása (R4.12., IV.4.1.).

F4.6: A védelmi szférában elsősorban a puha műveletek elhárítása terén hasznosítható azon hiányosság feltárása, hogy a polgárok titkolt érzelmei egyelőre jogilag nem megfoghatóak (egyelőre nem személyiségi jogok, mint pl. titkolt betegségeik), pedig ezekről az MI-t használva, akár visszaélési céllal is szerezhetőek adatok. (R1.3., I.4.)

- Ugyanez a meglátás jogvédelem számára is hasznos lehet.

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, itt nem összegzem őket, csupán néhány fontos és reális kutatás lehetőségét emelem ki közülük.

1. Először is: már folyamatban van a fejezetek szemelvényeiből egy egyetemi oktatásban használható formátumba való átdolgozás. Különböző fázisban készült munkapéldányokat teszt jelleggel már 2024 tavaszától a hallgatók számára rendelkezésre bocsátottam. A visszajelzések alapján tett javításokkal 2024 őszén immár egy teljesebb jegyzetet adtam közre tesztelésre, melyet 2025 tavaszára továbbfejlesztettem. A különböző kurzusok számára eltérő oldalszámú jegyzeteket tervezek, melyek alapja a II. fejezet, valamint néhány szemelvény a későbbi fejezetekből.
2. Kézenfekvő, az autonómiával kapcsolatos, egyelőre csupán sejtésként megfogalmazott gondolatkör komolyabb analízisének végrehajtása, melynek jelentőségét annak ismertetésekor bemutattam.
3. Hasznos lenne végigvinni azt a terjedelmi okokból kimaradt kutatást is, melyhez itt csupán meglátásként a vázlatát írtam le az „IQ-fenntarthatóság”-nak nevezett jelenségnek (IV.5.2.), bár nem az MI-hez kapcsolódik, hanem annál jóval tágabb problémát vet fel (itt ezért nem emeltem eddig ki). Úgy vélem, hogy ez a meglátás tudományosan is jól alátámasztható lesz a kutatás folytatása során.

4. Sürgető területnek tartom a kibertéri MI támadási és védelmi lehetőségeinek újravizsgálását, melyet érintettem ugyan (VI.2.1.), viszont egyrészt a témában elérhető több publikációt találtam, melyek feldolgozásra várnak, másrészt egyre jobban elérhetővé kezd válni az ilyen jellegű problémák gyakorlati vizsgálata is. A cél jellegéből adódóan ilyen tesztelésekhez nem feltétlenül szükségesek nagy teljesítményű hardverek, ami reálisabbá teszi az érdeklődést.
5. Szeretném a digitális erőérvényesítés-koncepció alkalmazását mielőbb megfogalmazni az oroszon 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.)

AJÁNLÁS – KIK SZÁMÁRA LEHET HASZNOS A FENTI TANULMÁNY?

Kedvcsinálónak itt kiemelek pár szakmát, ami nem jelenti, hogy az itt nem szereplők számára az anyag érdektelen. Sőt, remélem minden olvasóm talált érdekes, a maga számára hasznos anyagot. Bízom benne, hogy életükre is inspiratív hatással tudott lenni némely gondolat.

0. NEM kívánt a tanulmány az MI műszaki megoldásait kutató kollégáknak technikai újdonságokkal szolgálni, de nagyon jó lenne, ha minél több ilyen ember olvasná, hiszen az általuk elért eredmények itt elemzett aspektusai inspirációt adhatnak számukra is.
1. Elsődleges célközönség a **védelmi szféra szakemberei** voltak, a polgári és katonai, nemzetbiztonsági és rendészeti állományok egyaránt. Számukra a tanulmány számos helyén tett hasznosítható meglátásokat, de a VI. fejezet kifejezetten hasznos lehet. Közülük az MI-vel eddig nem sokat foglalkozók úgy nyerhetnek bevezetést ebbe a világba, hogy annak bemutatást hivatásuk fontos problémái illusztrálják, és az MI technológiai bemutatása kifejezetten az ilyen közönség számára lett megfogalmazva.
2. Ide kapcsolódó célközönség a **védelmi és katonai témák iránt érdeklődő** közönség, akik nem ilyen területen dolgoznak, de a számos sajátos megközelítés, modell, terminológia által mélyebben bepillantást nyerhetnek az erőérvényesítés kibontakozóban lévő tendenciáiba, mint egy-egy publikáció vagy népszerűsítő írás által.
3. A másik kiemelt célközönség az **oktatók**, elsősorban a felsőfokú intézményekben dolgozó kollégák. Ők már most is közvetlenül hasznosíthatnak számos alfejezetet, nem is kell meg-

várniuk, míg a megfelelő szemelvényekből jegyzet jellegű kivonat készül. Továbbá a katonai felsőoktatásról szóló rész egy, a védelmi és katonai felsőoktatás szakállománya felé irányított vitaindítóként is felfogható.

4. A **humán tudományok** szakemberei, illetve minden humán érdeklődésű ember számára a II. fejezet megközelítése lehet hasznos, hiszen feltehetőleg az 1. pontbeli állományokhoz hasonlóan csupán egy áttekintésre kíváncsiak az MI-ről, nem szeretnének mérnöki vagy programozói szinten elmélyülni a technológiákban. De számítok az ő érdeklődésükre a tanulmány filozófiai átitatottsága miatt is, mely sokféle ihletet adhat.
5. Jól felhasználhatják az anyagot a **szabályozásokban érintett szakemberek**, leginkább az informatikai biztonság kutatói és alkalmazói, de akár jogászok is. Számos felosztás és megközelítés irányult az MI-vel kapcsolatos szabályozási problémák feloldására, melyet saját szaktudásuk segítségével remélhetőleg továbbfejleszthetnek és a szabályozások megfogalmazási gyakorlatában alkalmazhatnak.
6. **Jogvédők** is jól használhatják a dolgozat egyes részeit, elsősorban a digitális visszaélések rendszerezését, melyre magyarul nem leltem fel ilyen átfogó tanulmányt, vagy pl. a polgárok titkolt érzelmeinek személyiségi jogokba való bevonásáról való meglátásomat.
7. Profitálhatnak a tanulmány gondolataiból a **gazdasági szakemberek** is. Elsősorban a többdimenziós fejlődési modellemet tartom e téren alkalmazhatónak, de az MI-regressziót felvető részekben és a védelmi fejezetben találhatnak számukra hasznosítható meglátásokat. Az etikus marketinghez hozzájárulhat a terminológiai degradáció modellje, a dolgozat egyéb vizsgálatai pedig hozzájárulhatnak prognózisaikhoz, vagy saját modelljeikhez.
8. A tudományok kimutatott konvergenciája miatt egészen biztos, hogy a tanulmány valamely részében bármilyen szakma találhat számára érdekes és hasznos részeket, vagyis remélem, hogy minden tudományág képviselője hasznosíthatja a kutatást, annak interdiszciplináris (és valamelyest multidiszciplináris) jellege miatt.

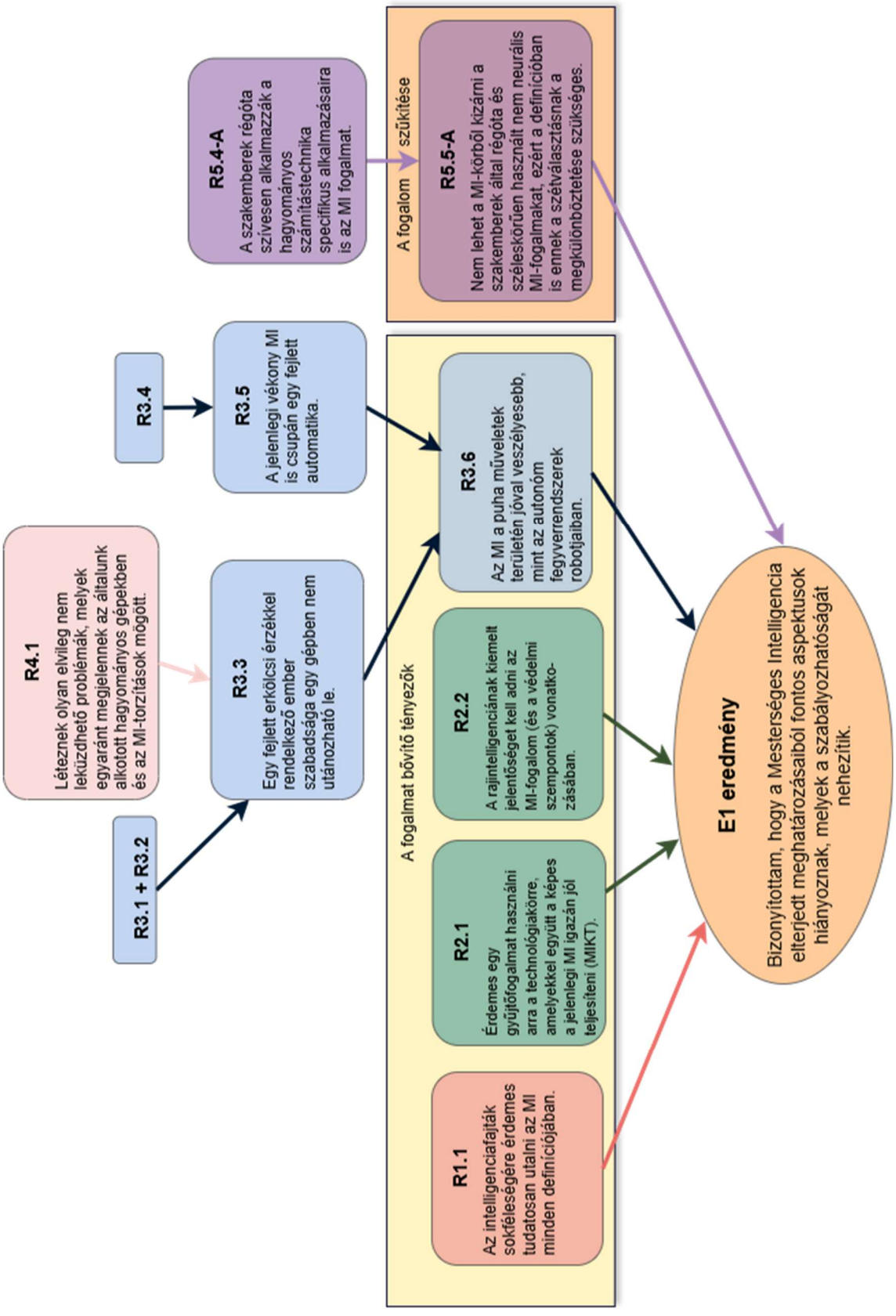
MUTATÓK, MELLÉKLETEK

ÁBRÁK JEGYZÉKE

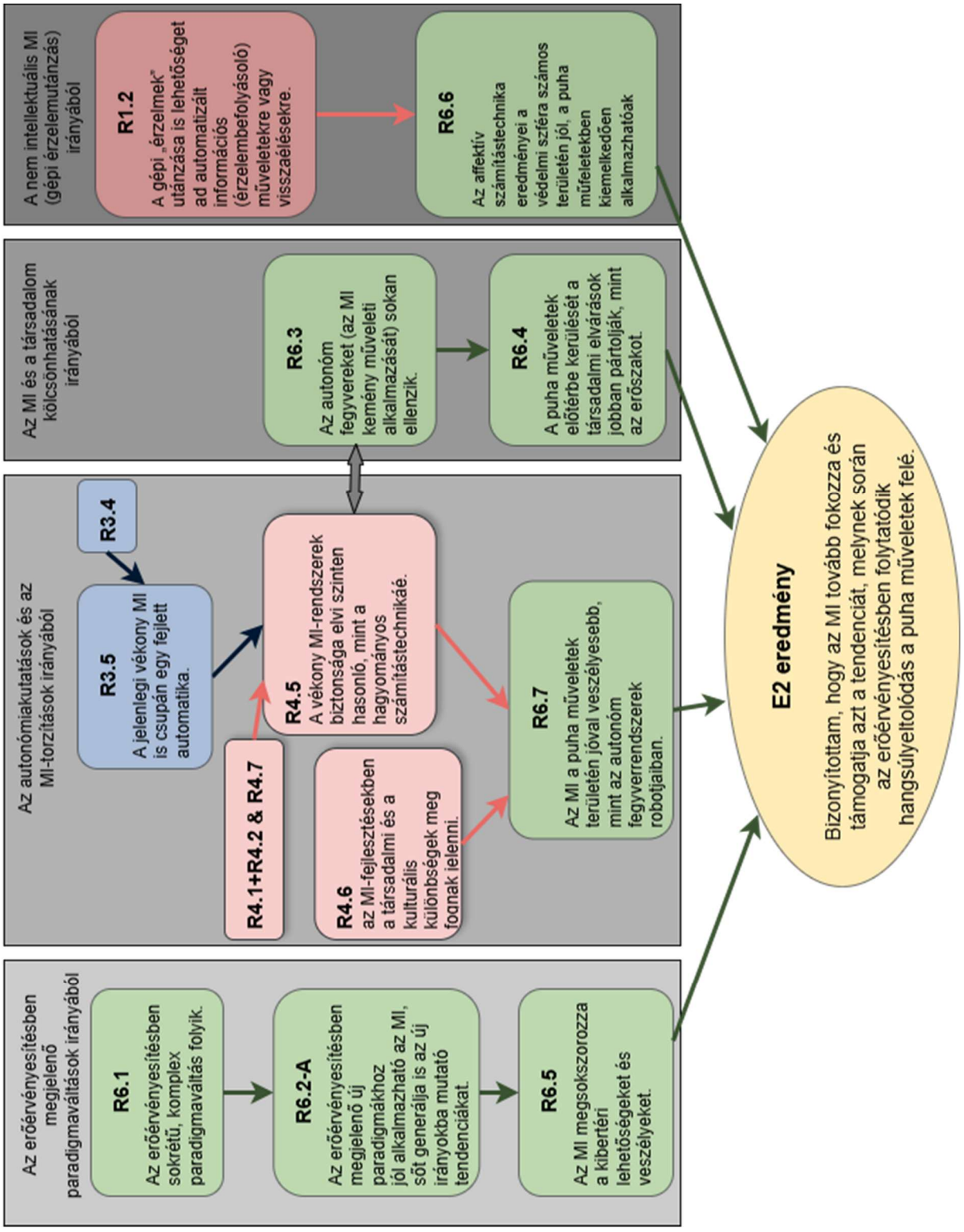
TÁBLÁZATOK JEGYZÉKE

KINAGYÍTOTT GONDOLATTÉRKÉPEK (OLVASHATÓBB VÁLTOZATOK)

21. ábra: A H1 bizonyításának gondolatrajza - nagyított változat (saját készítés)



22. ábra: A H2 bizonyításának gondolatrajza - nagyított változat (saját készítés)



IRODALOMJEGYZÉK

A megadott linkek kimásolva használhatóak, nem mindig működnek rájuk kattintva.

[1] Stuart Russell és mtsai.: *Mesterséges inteligencia elektronikus almanach*. Budapest: Panem Kft, 2019. Elérhető: https://dtk.tankonyvtar.hu/xmlui/handle/123456789/7622?show=full [Elérés: 2022. január 10.]

[2] Buzás György Miklós: *A mesterséges intelligencia története*, *Gasztroenterológiai és hepatológiai szemle*, köt. 7, sz. 3, Art. sz. 3, okt. 2021.

[3] Reding, Dale. F. & Jacqueline Eaton: *NATO ST Trends Report 2020-2040: Exploring the S&T Edge*. Belgium: NATO Science & Technology Organization, 2020.

[4] Oksana, Isabella: *Symbiotic Intelligence: Exploring the Intersection of Human and Artificial Intelligence*, *Best AI Digest*, 2023. október 27. Elérhető: https://medium.com/bestai/symbiotic-intelligence-exploring-the-intersection-of-human-and-artificial-intelligence-b2fd88a094d5. [Elérés: 2024. január 6.]

[5] Magyar Innovációs és Technológiai Minisztérium: *Magyarország Mesterséges Intelligencia Stratégiája 2020-2030*, Innovációs és Technológiai Minisztérium, Budapest, 2020. Elérhető: https://ai-hungary.com/api/v1/companies/15/files/137203/view. [Elérés: 2022. december 22.]

[6] EU: *Az első uniós rendelet a mesterséges intelligenciáról*, *Hírek*, 2023. június 12. Elérhető: https://www.europarl.europa.eu/topics/hu/article/20230601STO93804/az-elso-unios-rendelet-a-mesterseges-intelligenciarol. [Elérés: 2024. március 6.]

[7] Európa Parlament és Tanács rendelete: *A mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról*, 2023. június 14. Elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52021PC0206. [Elérés: 2023. november 3.]

[8] Bertuzzi, Luca: *EU lawmakers set to settle on OECD definition for Artificial Intelligence*, *www.euractiv.com*, 2023. március 7. Elérhető: https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/. [Elérés: 2023. november 6.]

[9] 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*, 2023. június 14. Elérhető: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0016.02/DOC_2 format=PDF. [Elérés: 2023. november 3.]

[10] Európa Bizottság MI Szakértői Csoport (szerk.): *Az AI (mesterséges intelligencia) meghatározása: Főbb képességek és tudományterületek.* Brüsszel: Publications Office of the EU, 2019. Elérhető: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60652 [Elérés: 2022. március 4.]

[11] Cominelli, Lorenzo Daniele Mazzei & Danilo Emilio Rossi: *SEAI: Social Emotional Artificial Intelligence Based on Damasio’s Theory of Mind*, *Front. Robot. AI*, köt. 5, pp. 6, 2018, doi: 10.3389/frobt.2018.00006

[12] Sayler, Kelley M.: *Artificial Intelligence and National Security*. Congressional Research SVC, 2020. Elérhető: https://crsreports.congress.gov/product/pdf/R/R45178. [Elérés: 2024. január 4.]

[13] Gosselin-Malo, Elisabeth: *NATO to update artificial intelligence strategy amid new threats*, *C4ISRNet*, 2023. november 30. Elérhető: https://www.c4isrnet.com/artificial-

intelligence/2023/11/30/nato-to-update-artificial-intelligence-strategy-amid-new-threats/. [Elérés: 2024. január 4.]
[14] Defense Innovation Board: *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense*. United States Department of Defense, 2019. Elérhető: https://media.defense.gov/2019/Oct/31/2002204459/-1/-1/0/DIB_AI_PRINCIPLES_SUPPORTING_DOCUMENT.PDF. [Elérés: 2023. december 18.]
[15] Fehér, András Tibor & Négyesi Imre: *A gépi érzelmek a fegyveres erőknél és az autonóm rendszerekben*, *Hadtudományi Szemle*, köt. 14, sz. 3, pp. 163–176, 2021, doi: 10.32563/hsz.2021.3.12
[16] Russell, Stuart J.Peter Norvig & John F. Canny: *Mesterséges intelligencia - modern megközelítésben*. Budapest: Panem, 2005.
[17] Hewitt, CarlPeter Bishop & Richard 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, pp. 235–245. Elérhető: https://www.ijcai.org/Proceedings/73/Papers/027B.pdf. [Elérés: 2024. szeptember 27.]
[18] Biddwan, Ahmed: *AI Agents - Types, Benefits and Examples*, *Yellow.ai*. Elérhető: https://yellow.ai/blog/ai-agents/. [Elérés: 2024. szeptember 28.]
[19] Cassell, Justine & Joseph SullivanScott Prevost & Elizabeth Churchill: *Embodied Conversational Agents*. MIT Press, 2000. Elérhető: https://mitpress.mit.edu/9780262032780/embodied-conversational-agents/ [Elérés: 2022. június 7.]
[20] Brooks, Rodney: *Just Calm Down About GPT-4 Already*, 2023. Elérhető: https://spectrum.ieee.org/gpt-4-calm-down. [Elérés: 2024. június 28.]
[21] Stone, Peter és mtsai.: *Artificial Intelligence and Life in 2030: The One Hundred Year Study on Artificial Intelligence*. arXiv, 2022. október 31. doi: 10.48550/arXiv.2211.06318. Elérhető: http://arxiv.org/abs/2211.06318. [Elérés: 2024. június 28.]
[22] Négyesi Imre: *A mesterséges intelligencia katonai felhasználásának lehetőségei*, köt. I. Budaepst: Zrinyi Kiadó, 2022.
[23] Gardner, Howard: *Frames of mind - The Theory of Multiple Intelligences*. New York: Basic Book, 1983. Elérhető: https://www.academia.edu/36707975/Frames_of_mind_the_theory_of_multiple_intelligences [Elérés: 2023. április 5.]
[24] Habeeb, Mahmood: *13 Types Of Intelligence (Which Ones Are You?) | LinkedIn*, *Capital Market Cafe*, 2023, Elérhető: https://www.linkedin.com/pulse/13-types-intelligence-which-ones-you-habeeb-mahmood/. [Elérés: 2024. január 18.]
[25] Vijay A., Kanade: *Narrow AI vs. General AI vs. Super AI*, *Spiceworks*, 2022. Elérhető: https://www.spiceworks.com/tech/artificial-intelligence/articles/narrow-general-super-ai-difference/. [Elérés: 2024. január 21.]
[26] Bostrom, Nick: *Szuperintelligencia [utak, veszélyek, stratégiák]*. Budapest: Ad Astra, 2015. Elérhető: https://pdfcoffee.com/nick-bostrom-szuperintelligencia-pdf-free.html [Elérés: 2024. július 9.]
[27] Pokol, Béla: *A mesterséges intelligencia: egy új létréteg kialakulása?*, *Informacios Tarsadalom*, köt. 17, sz. 4, pp. 39, 2017, doi: 10.22503/inftars.XVII.2017.4.3
[28] Vijay A., Kanade: *What Is Super Artificial Intelligence (AI)? Definition, Threats, and Trends*, *Spiceworks*. Elérhető: https://www.spiceworks.com/tech/artificial-intelligence/articles/super-artificial-intelligence/. [Elérés: 2024. január 31.]
[29] Kurzweil, Ray: *A szingularitás küszöbén*. Budapest: Ad Astra, 2013.

[30] Pokol Béla: *A mesterséges intelligencia társadalma*. Budapest: Kairosz, 2018.
[31] Waugh, Rob: *5 predictions in new book from Google futurist who foresaw the iPhone*, *Mail Online*, 2024. július 6. Elérhető: https://www.dailymail.co.uk/sciencetech/article-13596753/Tech-prophet-predicted-iPhone-years-advance-makes-alarming-forecasts-coming-years.html. [Elérés: 2024. július 8.]
[32] Tegmark, Max: *Élet 3.0*. HVG Kiadó Zrt, 2018.
[33] Bitera, Roland: *A kiterjsztett ember*, NKE HHK, Budapest, 2024.
[34] Pléh, Csaba: *A lélektan története. Bp: Osiris, 2010*. 2010. Elérhető: https://www.researchgate.net/publication/313656494_A_lelektan_tortenete_Bp_Osiris_2010. [Elérés: 2024. január 24.]
[35] Tomkins, Silvan S.: *Affect, imagery, consciousness*. New York: Springer, 1962.
[36] Tomkins, Silvan S.: *Interest—excitement.*, in *Affect, imagery, consciousness, Vol. 1: The positive affects.*, New York: Springer Publishing Co, 1992, pp. 336–368. doi: 10.1037/14351-010. Elérhető: http://content.apa.org/books/14351-010. [Elérés: 2024. január 24.]
[37] Csépe, ValériaMiklós Győri & Ragó Anett: *Általános pszichológia 1-3: 3. Nyelv, tudat, gondolkodás*. in Osiris tankönyvek. Budapest: Osiris, 2008. 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 [Elérés: 2022. március 9.]
[38] Nathanson, Donald L.: *Shame and pride: Affect, sex, and the birth of the self*. New York and London: Norton, 1992.
[39] Ekman, Paul: *Emotion in the human face*, Second edition. Malor books reprint Edition. Los Altos, California: Malor Books, 2013.
[40] Margitházi, Bea: *Érzelemgép, fertőzés és tükörneuronok*, *Metropolis*, 2021, Elérhető: http://metropolis.org.hu/erzelemgep-fertozes-es-tukorneuronok [Elérés: 2022. január 26.]
[41] Kun, Bernadette: *Az érzelmi intelligencia és az emocionális és szociális kompetenciák szerepe a pszichoaktív-szer-használatban*. Budaepst: ELTE doktori disszertáció, 2011. Elérhető: http://pszichologia.phd.elte.hu/vedesek/KunB_phd_disszertacioja.pdf. [Elérés: 2021. november 11.]
[42] Picard, Rosalind Wright: *Affective Computing*, *M.I.T Media Laboratory Perceptual Computing Section Technical Report*, köt. No. 321., 1995, Elérhető: https://affect.media.mit.edu/pdfs/95.picard.pdf [Elérés: 2022. február 13.]
[43] Turing, Alan M.: *Computing machinery and intelligence*, *Mind*, köt. LIX, sz. 236, pp. 433–460, 1950, doi: 10.1093/mind/LIX.236.433
[44] Murray, Iain R & John 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, pp. 1097–1108, 1993, doi: 10.1121/1.405558
[45] Lee, Jintae & T. L. Kunii: *Model-based analysis of hand posture*, *IEEE Computer Graphics and Applications*, köt. 15, sz. 5, pp. 77–86, 1995, doi: 10.1109/38.403831
[46] Pat Research: *What is Affective computing? Top 15 Affective computing Companies*, *PAT RESEARCH*, 2020, Elérhető: https://www.predictiveanalyticstoday.com/what-is-affective-computing/ [Elérés: 2022. szeptember 15.]
[47] Kunimatsu, Atsushi. és mtsai.: *Vector unit architecture for emotion synthesis*, *IEEE Micro*, köt. 20, sz. 2, pp. 40–47, 2000, doi: 10.1109/40.848471
[48] W. Picard, Rosalind & Tao Jianhua Tieniu Tan (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. Elérhető: https://link.springer.com/book/10.1007/11573548. [Elérés: 2023. január 4.]
[49] Minsky, Marvin: *The emotion machine: commensense thinking, artificial intelligence, and the future of the human mind*, 1. Simon & Schuster trade paperback ed. New York, Simon & Schuster, 2007.
[50] Bullington, Joseph (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
[51] Khashman, Adnan: *Emotional System for Military Target Identification Prof*, *NATO Research and Technology Organisation Symposium on Information Management-Exploitation, Stockholm*, 2009, Elérhető: https://www.semanticscholar.org/paper/Emotional-System-for-Military-Target-Identification-Khashman/598e392add0580da30ae50559e8361df27e34d0d. [Elérés: 2022. március 11.]
[52] Asada, Minoru: *Development of artificial empathy*, *Neuroscience research*, köt. 90, pp. 41–50, 2015, doi: 10.1016/j.neures.2014.12.002
[53] Zielinski, Radek: *Artificial intelligence being trained to show empathy*, *ReadWrite*, 2023. október 9. Elérhető: https://readwrite.com/artificial-intelligence-being-trained-to-show-empathy/. [Elérés: 2024. január 28.]
[54] EmoShape Confidental: *Emotion Processing Unit III Brochure*. EmoShape, 2018. Elérhető: https://emoshape.com/wp-content/uploads/2019/04/EPU-III-Brochure.pdf [Elérés: 2022. május 11.]
[55] Woods, Dan: *Why Big Data Needs Natural Language Generation to Work*, Elérhető: https://www.forbes.com/sites/danwoods/2015/07/09/why-big-data-needs-natural-language-generation-to-work/?sh=7604e07b156c. [Elérés: 2021. január 27.]
[56] Fehér, András Tibor: *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*, Dr. Magyar, Sándor & Dr. Kenedli Tamás (szerk.), Budapest: KNBSZ, 2025., pp. 173-209. https://www.knbsz.gov.hu/kiadvanyok#1fd723c5-92d3-4dff-bc47-3edf4cb1b38a [Elérés: 2025. július 12.]
[57] *Big data: definition, benefits, challenges (infographics) | News | European Parliament*, 2021. február 17. Elérhető: https://www.europarl.europa.eu/news/en/headlines/society/20210211STO97614/big-data-definition-benefits-challenges-infographics. [Elérés: 2023. november 2.]
[58] Network of the National Library of Medicine: *Data Lake definition*, *NNLM*, 2024. Elérhető: https://www.nnlm.gov/guides/data-glossary/data-lake. [Elérés: 2024. július 14.]
[59] Szűts, Zoltán & Jinil Yoo: *Big Data, az információs társadalom új paradigmája*, *InfTars*, köt. 16, sz. 1, pp. 8, márc. 2016, doi: 10.22503/inftars.XVI.2016.1.1
[60] Farkas, Luca Ágnes & Alexandra Vida: *Big Data az erőfölény vonatkozásában (a digitális platformok meghódítója vagy a versenyhatóságok legnagyobb fejtörése)*, *a Gazdasági és Verseny Hivatal pályázatán nyertes tanulmány*, 2020, Elérhető: https://www.gvh.hu/pfile/file?path=/gvh/versenykultura_fejlesztes/tanulmanyi_verseny/palyazati_eredmenyek/farkas-luca-agnes---vida-alexandra&inline=true [Elérés: 2023. április 16.]
[61] Mell, Peter & Tim 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. Elérhető: https://csrc.nist.gov/pubs/sp/800/145/final. [Elérés: 2023. november 3.]
[62] Vadász, Dénes: *Operációs rendszerek*, 2002. kiad. Miskolc: Miskolci Egyetem, 2002. Elérhető: https://users.iit.uni-miskolc.hu/~vadasz/GEIAL302B/GEIAL202-Operacios-rendszerek-jegyzet.pdf [Elérés: 2022. január 25.]
[63] Aliyu, Ahmed és mtsai.: *Cloud Computing in VANETs: Architecture, Taxonomy, and Challenges*, *IETE Technical Review*, köt. 35, sz. 5, pp. 523–547, 2018, doi: 10.1080/02564602.2017.1342572
[64] Antalóczy Tibor: *Így működik a Tesla táblafelismerő rendszere*, *Villanyautósok*, 2020. április 27. Elérhető: https://villanyautosok.hu/2020/04/27/igy-mukodik-a-tesla-tablafelismero-rendszere/. [Elérés: 2024. március 4.]
[65] *Internet of Things (IoT)*, *ENISA*. Elérhető: https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/iot. [Elérés: 2023. november 2.]
[66] Török, Péter & Imre Négyesi: *A short overview for the cognition of the Internet of Things (MTMT)*, *RED - American Journal Of Research Education And Development*, sz. 2, pp. 4–15, 2017.
[67] Malin, Carlberg és mtsai: *Industry 4.0.*, Directorate-General for Internal Policies of the Union (European Parliament), LU Publications Office of the European Union, 2016. Elérhető: https://data.europa.eu/doi/10.2861/947880. [Elérés: 2023. november 2.]
[68] Bery, Nandita: *Security Threats in Smart Factories and Industry 4.0*, *Interconnections - The Equinix Blog*, 2023. március 14. Elérhető: https://blog.equinix.com/blog/2023/03/14/security-threats-in-smart-factories-and-industry-4-0/. [Elérés: 2024. február 24.]
[69] Dürr, OliverBeate Sick& Elvis Murina: *Probabilistic deep learning: with Python, Keras and TensorFlow Probability*. in Exercises in Jupyter Notebooks. Shelter Island: Manning, 2020. Elérhető: https://freecontent.manning.com/neural-network-architectures/. [Elérés: 2024. május 15.]
[70] Chowdhury, Prasenjit: *Perceptron: A Simple yet Mighty Machine Learning Algorithm*, *Medium*, 2023. június 19. Elérhető: https://medium.com/@cprasenjit32/perceptron-a-simple-yet-mighty-machine-learning-algorithm-9ff6b7d86a71. [Elérés: 2024. január 13.]
[71] Dudyrev, EgorIlia és mtsai: *Human knowledge models: Learning applied knowledge from the data*, *PLOS ONE*, köt. 17, sz. 10, pp. e0275814, okt. 2022, doi: 10.1371/journal.pone.0275814
[72] Yu, Xiang Qinghua & Zhou Shuihua Wang & Yudong Zhang: *A systematic survey of deep learning in breast cancer*, *International Journal of Intelligent Systems*, köt. 37, aug. 2021, doi: 10.1002/int.22622. Elérhető: https://www.researchgate.net/publication/353953952_A_systematic_survey_of_deep_learning_in_breast_cancer. [Elérés: 2024. január 13.]
[73] Shiri, Farhad és mtsai: *A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU*, *arXiv.org*, 2023. május 27. Elérhető: https://arxiv.org/abs/2305.17473v2. [Elérés: 2024. július 24.]
[74] Chakure, Afroz: *Convolutional Neural Networks (CNN) in a Brief*, *DEV Community*, 2021. január 20. Elérhető: https://dev.to/afrozchakure/cnn-in-a-brief-27gg. [Elérés: 2024. április 30.]
[75] Nelson, Daniel: *Mik azok a CNN-ek (konvolúciós neurális hálózatok)? - Egyesüljetek.AI*, *UNITE AI*, 2020. Elérhető: https://www.unite.ai/hu/mik-azok-a-konvol%C3%BAci%C3%B3s-neur%C3%A1lis-h%C3%A1l%C3%B3zatok/. [Elérés: 2024. április 30.]

[76] Fazekas, László: *TensorFlow alapozó*, *Medium*, 2019. december 19. Elérhető: https://thebojda.medium.com/tensorflow-alapoz%C3%B3-d2d1ee97c9db. [Elérés: 2024. április 30.]
[77] Gogul, I. & Sathiesh Kumar: *Flower species recognition system using convolution neural networks and transfer learning*, *International Conference on Signal Processing, Communications and Networking*, pp. 1–6, márc. 2017, doi: 10.1109/ICSCN.2017.8085675
[78] Bhat, Harshini: *Recurrent Neural Network: Applications and Advancements*, *AlmaBetter*, 2023. augusztus 1. Elérhető: https://www.almabetter.com/bytes/articles/recurrent-neural-network. [Elérés: 2024. január 14.]
[79] Rácz Bianka & Csapó Tamás Gábor: *Ajakvideó alapú beszédszintézis konvolúciós és rekurrens mély neurális hálózatokkal*, *Beszédtudomány - Speech Science*, köt. 1, pp. 57–72, dec. 2020, doi: 10.15775/Besztud.2020.57-72
[80] Prodip, Hore & Chatterjee Sayan: *A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone*, *Analytics Vidhya*, 2019. november 20. Elérhető: https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/. [Elérés: 2024. január 14.]
[81] Bahdanau, Dzmitry & Kyunghyun Cho & Yoshua Bengio: *Neural Machine Translation by Jointly Learning to Align and Translate*. arXiv, 2016. május 19. doi: 10.48550/arXiv.1409.0473. Elérhető: http://arxiv.org/abs/1409.0473. [Elérés: 2024. május 6.]
[82] Lippe, Phillip: *6. oktatóanyag: Transzformátorok és többfejes figyelem – UVA DL notebookok v1.2 dokumentációja*, in *UvA Deep Learning Tutorials*, UvA.2022. Elérhető: https://uvadlc-notebooks.readthedocs.io/en/latest/tutorial_notebooks/tutorial6/Transformers_and_MHAttention.html. [Elérés: 2024. május 7.]
[83] Inuwa, Mobarak: *Understanding Attention Mechanisms Using Multi-Head Attention*, *Analytics Vidhya*, 2023. június 22. Elérhető: https://www.analyticsvidhya.com/blog/2023/06/understanding-attention-mechanisms-using-multi-head-attention/. [Elérés: 2024. május 7.]
[84] Vaswani, Ashish és mtsai.: *Attention Is All You Need*, *31st Conference on Neural Information Processing Systems, Long Beach*, aug. 2023, doi: 10.48550/arXiv.1706.03762. Elérhető: http://arxiv.org/abs/1706.03762. [Elérés: 2024. január 14.]
[85] Kesrwani, Akash: *Multi-Head Self Attention: Short Understanding*, *Medium*, 2023. szeptember 8. Elérhető: https://medium.com/@akash.kesrwani99/multi-head-self-attention-short-understanding-e90a34866730. [Elérés: 2024. január 14.]
[86] Daniel, Nelson: *What is a Generative Adversarial Network (GAN)?*, *Unite AI*, 0 2020, Elérhető: https://www.unite.ai/what-is-a-generative-adversarial-network-gan/. [Elérés: 2024. január 14.]
[87] Muhamedyev, Ravil I: *Machine learning methods: An overview*, *Computer Modelling & New Technologies*, sz. 19, pp. 14–29, 2015.
[88] Sajid, Haziga: *Gépi tanulás vs. mély tanulás – Főbb különbségek – Unite.AI*, 2023. Elérhető: https://www.unite.ai/machine-learning-vs-deep-learning-key-differences/. [Elérés: 2024. január 10.]
[89] Ovtcharov, Kalin és mtsai: *Accelerating Deep Convolutional Neural Networks Using Specialized Hardware*, febr. 2015, Elérhető: https://www.microsoft.com/en-us/research/publication/accelerating-deep-convolutional-neural-networks-using-specialized-hardware/. [Elérés: 2024. január 8.]

[90] Schneider, Josh: *FPGA vs. GPU for Deep Learning Applications | IBM*, *IBM Blog*, 2024. május 10. Elérhető: https://www.ibm.com/think/topics/fpga-vs-gpu. [Elérés: 2025. január 8.]
[91] Kousi, Helena: *GPU, LPU and NPU: What are these architectures? - DataNorth*, *https://datanorth.ai*, 2024. november 4. Elérhető: https://datanorth.ai/blog/gpu-lpu-npu-architectures. [Elérés: 2025. január 2.]
[92] Asztalos, Olivér: *AI-ra gyúr a legújabb Huawei Kirin processzor*, *HWSW*, 2017. Elérhető: https://www.hwsw.hu/hirek/57742/huawei-hisilicon-kirin-970-processzor-soc-rendszerchip-npu-okostelefon-mi-ai.html. [Elérés: 2024. január 2.]
[93] Huawei, Atlas 800 Server: *NPU Board Components - NPU Board Components - Overview*, 2024. Elérhető: https://support.huawei.com/enterprise/en/doc/EDOC1100149965/d0bd4bfe/npu-board-components. [Elérés: 2025. január 2.]
[94] Kennedy, Patrick: *SambaNova SN10 RDU at Hot Chips 33*, *ServeTheHome*, 2021. augusztus 24. Elérhető: https://www.servethehome.com/sambanova-sn10-rdu-at-hot-chips-33/. [Elérés: 2025. január 2.]
[95] Davies, Mike: *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, Elérhető: https://download.intel.com/newsroom/2021/new-technologies/intel-labs-loihi-2-lava-quotes.pdf. [Elérés: 2024. január 6.]
[96] Zhang, Hai-Tian és mtsai.: *Reconfigurable perovskite nickelate electronics for artificial intelligence*, *Science (New York, N.Y.)*, köt. 375, sz. 6580, pp. 533–539, 2022, doi: 10.1126/science.abj7943
[97] Dienes, István & Stratégiai Kutató Intézet: *Mesterséges intelligencia másképpen: kvantumszámítógépek és a topologikus energia*, *ÁT - Áram és Technológia*, sz. III. évfolyam 1. szám, 2005, Elérhető: http://www.metaelmelet.hu/pdfek/mesterseges_intelligencia_maskeppen.pdf [Elérés: 2024. május 10.]
[98] Peral-García, David és mtsai: *Systematic literature review: Quantum machine learning and its applications*, *Computer Science Review*, köt. 51, pp. 100619, febr. 2024, doi: 10.1016/j.cosrev.2024.100619
[99] Jerbi, Sofiene és mtsai: *Shadows of quantum machine learning*, *Nat Commun*, köt. 15, sz. 1, pp. 5676, júl. 2024, doi: 10.1038/s41467-024-49877-8
[100] Swayne, Matt: *2025 Expert Quantum Predictions — Quantum Computing*, *The Quantum Insider*, 2024. december 31. Elérhető: https://thequantuminsider.com/2024/12/31/2025-expert-quantum-predictions-quantum-computing/. [Elérés: 2025. január 8.]
[101] Park, A.B. Wu & 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, pp. 89–110, 1998, doi: 10.1163/156856298x00451
[102] Smirnova, Lena és mtsai: *Brain-Cell Cultures: The Future of Computers and More?*, *Frontiers for Young Minds*. Elérhető: https://kids.frontiersin.org/articles/10.3389/frym.2023.1049593. [Elérés: 2024. február 25.]
[103] Bennet, Devasier & Tuan Vo-Dinh & Frederic Zenhausern: *Current and emerging opportunities in biological medium-based computing and digital data storage*, *Nano Select*, köt. 3, sz. 5, pp. 883–902, 2022, doi: 10.1002/nano.202100275

[104] Hamzelou, Jessica: *A memory prosthesis could restore memory in people with damaged brains*, *MIT Technology Review*, 2022. Elérhető: https://www.technologyreview.com/2022/09/06/1059032/memory-prosthesis-damaged-brains/. [Elérés: 2024. február 25.]
[105] Oláh, Ferenc: *A fuzzy logika - alapismeretek*, *autotechnika.hu*, 2009. szeptember 7. Elérhető: https://autotechnika.hu/cikkek/motor-eroatvitel/8313/a-fuzzy-logika-alapismeretek. [Elérés: 2024. május 9.]
[106] Zadeh, Lotfi Aliasker: *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, pp. 28–44, 1973, doi: 10.1109/TSMC.1973.5408575
[107] MATLAP center: *Defuzzification Methods*. Elérhető: https://www.mathworks.com/help/fuzzy/defuzzification-methods.html. [Elérés: 2024. december 9.]
[108] Zadeh, Lotfi Aliasker: *Fuzzy sets*, *Information and Control*, köt. 8, sz. 3, pp. 338–353, jún. 1965, doi: 10.1016/S0019-9958(65)90241-X
[109] Garrido, Angel: *A Brief History of Fuzzy Logic*, *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, köt. 3, sz. 1, Art. sz. 1, 2012.
[110] Pease, Bob: *What's All This Fuzzy Logic Stuff, Anyhow?*, *Electronic Design*, 1993. május. Elérhető: https://www.electronicdesign.com/technologies/embedded/digital-ics/article/21757343/whats-all-this-fuzzy-logic-stuff-anyhow. [Elérés: 2024. december 9.]
[111] Cintula, Petr & Christian G. Fermüller & Carles Noguera: *Fuzzy Logic*, in *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. & U. Nodelman, (szerk.), Summer 2023.Metaphysics Research Lab, Stanford University, 2023. Elérhető: https://plato.stanford.edu/archives/sum2023/entries/logic-fuzzy/. [Elérés: 2024. december 9.]
[112] Simran Kaur, Arora: *What is Fuzzy Logic? Advantages and Disadvantages*, *Hackr.io*. Elérhető: https://hackr.io/blog/what-is-fuzzy-logic. [Elérés: 2024. december 10.]
[113] Otten, Neri Van: *Fuzzy Logic Made Easy — Its Application In AI, Machine Learning & Natural Language Processing (NLP)*, *Spot Intelligence*, 2023. február 13. Elérhető: https://spotintelligence.com/2023/02/13/fuzzy-logic-ai-machine-learning/. [Elérés: 2024. december 10.]
[114] Global, iCert: *What is Fuzzy Logic in AI and its uses?*, *icertglobal.com*. Elérhető: https://www.icertglobal.com/what-is-fuzzy-logic-in-ai-and-its-uses-blog/detail. [Elérés: 2024. december 10.]
[115] Benavides, Efrén: *Fuzzy Ethics: A moral criterion for sustainability*. 2013.
[116] PentaSchool Oktatási Központ: *A Neumann elv*, *Világhíres Feltalálóink*. Elérhető: http://www.feltalaloink.hu/tudosok/neumannjanos/html/neujantal2.htm. [Elérés: 2024. január 16.]
[117] Neumann, János: *A számológép és az agy*. Elérhető: https://mek.oszk.hu/01200/01255/html/#02. [Elérés: 2024. január 16.] [Elérés: 2024. január 6.]
[118] *Bionics*, *Enciclopedia Britannica*. 2022. Elérhető: https://www.britannica.com/technology/bionics. [Elérés: 2024. január 16.]
[119] Darwish, Ashraf: *Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications*, *Future Computing and Informatics Journal*, köt. 3, sz. 2, pp. 231–246, dec. 2018, doi: 10.1016/j.fcij.2018.06.001
[120] Dorigo, Marco & Mauro Birattari & Thomas Stützle: *Ant Colony Optimization*, *Computational Intelligence Magazine, IEEE*, köt. 1, pp. 28–39, dec. 2006, doi: 10.1109/MCI.2006.329691

[121] Zelinka, Ivan és mtsai: *Swarm virus - Next-generation virus and antivirus paradigm?*, *Swarm and Evolutionary Computation*, köt. 43, pp. 207–224, dec. 2018, doi: 10.1016/j.swevo.2018.05.003
[122] Karaboga, Dervis & Bahriye Akay: *Akay, B.: A Survey: Algorithms Simulating Bee Swarm Intelligence. Artificial Intelligence Review 31, 68-85*, *Artif. Intell. Rev.*, köt. 31, pp. 61–85, jún. 2009, doi: 10.1007/s10462-009-9127-4
[123] Lin, Kuan-Cheng & Sih-Yang Chen & Jason Hung: *Botnet Detection Using Support Vector Machines with Artificial Fish Swarm Algorithm*, *Journal of Applied Mathematics*, köt. 2014, pp. 1–9, ápr. 2014, doi: 10.1155/2014/986428
[124] Fister, Iztok és mtsai: *A comprehensive review of firefly algorithms*, *Swarm and Evolutionary Computation*, köt. 13, pp. 34–46, dec. 2013, doi: 10.1016/j.swevo.2013.06.001
[125] Bohre, Aashish & Ganga Agnihotri & Manisha 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] Yang, Xin-She: *Nature-inspired Metaheuristic Algorithms*, 2. kiadás. Luniver Press, 2010. Elérhető: https://staff.fmi.uvt.ro/~daniela.zaharie/ma2016/projects/techniques/FireflyAlgorithm/Yang_nature_book_part.pdf [Elérés: 2023. március 14.]
[127] Faris, H.I. Aljarah & M. Al-Betar & S. Mirjalili: *Grey wolf optimizer: a review of recent variants and applications*, *Neural Computing and Applications*, köt. 10, pp. 1–24, 2017.
[128] Panimalar S., Arockia: *Nature Inspired Metaheuristic Algorithms*, *International Research Journal of Engineering and Technology (IRJET)*, 2017. Elérhető: https://www.irjet.net/archives/V4/i10/IRJET-V4I1054.pdf. [Elérés: 2024. május 10.]
[129] Du, Zhi-Gang és mtsai: *Improved Binary Symbiotic Organism Search Algorithm With Transfer Functions for Feature Selection*, *IEEE Access*, köt. 8, pp. 225730–225744, 2020, doi: 10.1109/ACCESS.2020.3045043
[130] Kwa, Hian & mtsai.: *Adaptivity: a path towards general swarm intelligence?*, *Frontiers in Robotics and AI*, köt. 10, 2023, doi: 10.3389/frobt.2023.1163185. Elérhető: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10203170/. [Elérés: 2024. január 22.]
[131] Fehér, András Tibor & Kereszthegyi István: *A gépi nyelvfeldolgozás védelmi felhasználási lehetőségei és kockázata*, *Szakmai Szemle*, köt. XXI., sz. 4., pp. 75–93, 0 2023.
[132] Microsoft ügyfélszolgálat: *A Microsoft 365 Copilot áttekintése*, 2023. november 12. Elérhető: https://learn.microsoft.com/hu-hu/microsoft-365-copilot/microsoft-365-copilot-overview. [Elérés: 2023. november 16.]
[133] Chowdhary, K. R.: *Natural Language Processing*, in *Fundamentals of Artificial Intelligence*, Chowdhary, K. R., (szerk.), New Delhi, Springer India, 2020, pp. 603–649. doi: 10.1007/978-81-322-3972-7_19. Elérhető: https://doi.org/10.1007/978-81-322-3972-7_19. [Elérés: 2023. november 13.]
[134] Váradi, Tamás és mtsai.: *Az e-magyar digitális nyelvfeldolgozó rendszer*, Vincze V., Szegedi Tudományegyetem Informatikai Tanszékcsoport, 2017, pp. 49–60. Elérhető: http://real.mtak.hu/72361/. [Elérés: 2023. november 24.]
[135] Kumar, Gautam: *Role of Tokenization in NLP*, *Medium*, 2023. augusztus 10. Elérhető: https://kmr-gautam2893.medium.com/role-of-tokenization-in-nlp-18057618a102. [Elérés: 2023. november 24.]
[136] Shrivastava, Anshumali: *Key-Value Databases are Sufficient Infrastructure for Semantic Search at Scale with NeuralDB.*, *ThirdAI Blog*, 2024. február 10. Elérhető:

https://medium.com/thirdai-blog/key-value-databases-are-sufficient-infrastructure-for-semantic-search-at-scale-with-neuraldb-b7eea4b9b4db. [Elérés: 2024. március 12.]
[137] Darwall, Stephen: *The Value of Autonomy and Autonomy of the Will**, *Ethics*, köt. 116, pp. 263–284, 2006.
[138] Váróné Tomori Viola: *Kurt Lewin és a mezőelmélet*, *Korunk*, sz. 32. évf. 2. sz. (1973. február), pp. 312–316, 1973.
[139] Endsley, Mica R. & David B. Kaber: *Level of automation effects on performance, situation awareness and workload in a dynamic control task*, *Ergonomics*, köt. 42, sz. 3, pp. 462–492, márc. 1999, doi: 10.1080/001401399185595
[140] J3016 szabvány (2021.04.): *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*. Elérhető: https://www.sae.org/standards/content/j3016_202104/. [Elérés: 2024. július 12.]
[141] Simon, Zsolt: *Ezek az önvezetés szintjei*, *Villanyautósok*, 2021. május 11. Elérhető: https://villanyautosok.hu/2021/05/11/ezek-az-onvezetes-szintjei/. [Elérés: 2024. július 12.]
[142] C&T: *The 6 Levels of Autonomous Driving Explained*, *C&T Solution Inc.* Elérhető: https://www.candtsolution.com/news_events-detail/the-six-level-of-autonomous-driving-explained/. [Elérés: 2024. július 12.]
[143] Beer, Jenay & Arthur Fisk & Wendy Rogers: *Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction*, *Journal of Human-Robot Interaction*, köt. 3, pp. 74, jún. 2014, doi: 10.5898/JHRI.3.2.Beer
[144] Krä, Michaela és mtsai: *Levels of autonomy in production logistics: terminology and framework*, 2022, Elérhető: https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/96145. [Elérés: 2024. július 12.]
[145] Attanasio, Aleks & mtsai.: *Autonomy in Surgical Robotics*, *Annual Review of Control, Robotics, and Autonomous Systems*, köt. 4, sz. Volume 4, 2021, pp. 651–679, máj. 2021, doi: 10.1146/annurev-control-062420-090543
[146] Papdi-Pécskői, Viktor: *Emberek vezették a Cruise robotaxijait*, 2023. november 12. Elérhető: https://index.hu/techtud/2023/11/12/robotaxi-onvezeto-egyesult-allamok-san-francisco-general-motors-cruise-origin-felfuggesztes-godrok-gyerekek/. [Elérés: 2024. augusztus 27.]
[147] Asimov, Isaac: *A két évszázados ember*, in: Robottörténetek 2., Móra, 1993.
[148] Heller, Ágnes: *Általános etika*. Budapest, Cserépfalvi, 1994.
[149] IBM Support: *Maximo for Civil Infrastructure 7.6.2*, 2022. május 11. Elérhető: https://www.ibm.com/docs/hu/mfci/7.6.2?topic=implementing-highly-available-systems. [Elérés: 2024. augusztus 21.]
[150] Coolidge, Matt: *The Danger of AI's Black Box | Casper Labs*, 2023. július 6. Elérhető: https://casperlabs.io/blog/the-danger-of-ais-black-box. [Elérés: 2024. október 17.]
[151] Bhatt, Umang 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, pp. 401–413. doi: 10.1145/3461702.3462571. Elérhető: https://dl.acm.org/doi/10.1145/3461702.3462571. [Elérés: 2024. november 29.]
[152] Stogiannos, Nikolaos 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, pp. 105423, jún. 2024, doi: 10.1016/j.ijmedinf.2024.105423

[153] Balogh, Nóra: *XAI, avagy a magyarázható mesterséges intelligencia*, *Dmlab*, 2021. Elérhető: https://dmlab.hu/blog/xai-avagy-a-magyarazhato-mesterseges-intelligencia/. [Elérés: 2024. október 17.]
[154] Goodfellow, Ian J. & Jonathon Shlens & Christian Szegedy: *Explaining and Harnessing Adversarial Examples*. arXiv, 2015. március 20. doi: 10.48550/arXiv.1412.6572. Elérhető: http://arxiv.org/abs/1412.6572. [Elérés: 2024. október 16.]
[155] Kansky, Ken és mtsai.: *Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics*. arXiv, 2017. augusztus 17. doi: 10.48550/arXiv.1706.04317. Elérhető: http://arxiv.org/abs/1706.04317. [Elérés: 2024. október 17.]
[156] Chollet, François: *Deep learning with Python*. Shelter Island, NY: Manning, 2018. Elérhető: https://blog.keras.io/the-limitations-of-deep-learning.html
[157] Webber, Francisco: *Third AI Winter ahead? Why OpenAI, Google & Co are heading towards a dead-end | Cortical.io*, 2022. július 18. Elérhető: https://www.cortical.io/blog/third-ai-winter-ahead-why-openai-google-co-are-heading-towards-a-dead-end/. [Elérés: 2024. október 27.]
[158] Lee, Nicol Turner & Paul Resnick & Genie Barton: *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, Brookings, máj. 2019. Elérhető: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/. [Elérés: 2022. december 6.]
[159] Matyszczyk, Chris: *Thanks, Twitter. You turned Microsoft's AI teen into a horny racist*, *CNET*. Elérhető: https://www.cnet.com/culture/twitter-turned-microsoft-ai-teen-tay-into-horny-foul-mouthed-sex-bot-racist/. [Elérés: 2025. június 29.]
[160] Dervenkár István: *Előítélet-detektálót csinált az IBM mesterséges intelligenciához*, *Bitport*, 2018. Elérhető: https://bitport.hu/eloitelet-detektalot-csinalt-az-ibm-mesterseges-intelligenciahoz/. [Elérés: 2022. december 18.]
[161] Manyika, James & Jake Silberg: *Tackling bias in artificial intelligence (and in humans)*. McKinsey Global Institute, 2019. Elérhető: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans [Elérés: 2023. április 16.]
[162] Fehér, András Tibor: *A mesterségesintelligencia-alapú hidegháború etikai háttere*, *A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata*, pp. 355–392, okt. 2023.
[163] Molnár, Géza Gábor: *Szemantikus web technológiák alkalmazási lehetőségei az „exploratory" OLAP-ban*, phd, Budapesti Corvinus Egyetem, 2022. Elérhető: https://doi.org/10.14267/phd.2022062. [Elérés: 2024. július 9.]
[164] Gontier, Elodie Marie: *Web Semantic and Ontology*, *AIT*, köt. 05, sz. 02, pp. 15–20, 2015, doi: 10.4236/ait.2015.52003
[165] Alexander Serov: *Subjective Reality and Strong Artificial Intelligence*. 2013. Elérhető: https://www.researchgate.net/publication/235221450_Subjective_Reality_and_Strong_Artificial_Intelligence [Elérés: 2022. május 17.]
[166] Koebler, Jason: *We Asked MIT Researchers Why They Made a 'Psychotic AI' That Only Sees Death*, *Vice*, 2018. június 7. Elérhető: https://www.vice.com/en/article/xwm5mk/mit-psychotic-ai-rehabilitation. [Elérés: 2022. december 18.]
[167] Salih, Ameer M. és mtsai.: *Assessment of Chat-GPT, Gemini, and Perplexity in Principle of Research Publication: A Comparative Study*, *Barw Medical Journal*, 2025, doi: 10.58742/bmj.v2i4.140. Elérhető:

https://www.barwmedical.com/index.php/BMJ/article/view/140. [Elérés: 2025. június 20.]
[168] Xian, Yongqin & Bernt Schiele & Zeynep Akata: *Zero-Shot Learning -- The Good, the Bad and the Ugly*. arXiv, 2020. szeptember 23. doi: 10.48550/arXiv.1703.04394. Elérhető: http://arxiv.org/abs/1703.04394. [Elérés: 2025. június 29.]
[169] *Prevision.io — Documentation — Documentation PrevisionIO*. Elérhető: https://previsionio.readthedocs.io/fr/latest/index.html. [Elérés: 2022. december 18.]
[170] Vaithianathan, Rhema & Emily Kulick & Diana Benavides Prado: *Allegheny Family Screening Tool: Methodology, Version 2 - Centre for Social Data Analytics - AUT*, Centre for Social Data Analytics, 2019. Elérhető: https://csda.aut.ac.nz/research/our-publications/2019/allegheny-family-screening-tool-methodology,-version-2. [Elérés: 2022. december 18.]
[171] Marr, Bernard: *How To Solve AI's Bias Problem, Create Emotional AIs, And Democratize AI With Synthetic Data*, *Forbes*. 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/. [Elérés: 2022. december 20.]
[172] Peychev, Momchil & mtsai.: *Latent Space Smoothing for Individually Fair Representations*. arXiv, 2022. július 26. Elérhető: http://arxiv.org/abs/2111.13650. [Elérés: 2022. december 20.]
[173] Drage, Eleanor & Kerry Mackereth: *Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference"*, *Philos. Technol.*, köt. 35, sz. 4, pp. 89, okt. 2022, doi: 10.1007/s13347-022-00543-1
[174] Mohanani, Rahul & mtsai.: *Cognitive Biases in Software Engineering: A Systematic Mapping Study*, *IEEE Transactions on Software Engineering*, köt. 46, sz. 12, pp. 1318–1339, 2020, doi: 10.1109/TSE.2018.2877759
[175] Fehér, András Tibor: *Mesterséges intelligencia a kibervédelemben*, in *A hadtudomány aktuális kérdései napjainkban*, Szelei Ildikó., (szerk.), Budapest: Ludovika Egyetemi Kiadó, 2022, pp. 123–138. Elérhető: https://tudasportal.uni-nke.hu/xmlui/handle/20.500.12944/17628. [Elérés: 2022. december 7.]
[176] Zhou, Yuhan & mtsai.: *A Survey on Data Quality Dimensions and Tools for Machine Learning Invited Paper*, in *2024 IEEE International Conference on Artificial Intelligence Testing (AITest)*, júl. 2024, pp. 120–131. doi: 10.1109/AITest62860.2024.00023. Elérhető: https://ieeexplore.ieee.org/document/10685186. [Elérés: 2025. július 10.]
[177] Golubovskij, Vladislav: *Explainability and Transparency of AI in Auditing*, 2025. június 26. Elérhető: http://essay.utwente.nl/106877/. [Elérés: 2025. július 10.]
[178] Roshinta, Trisna Ari & Szűcs Gábor: *A Comparative Study of LIME and SHAP for Enhancing Trustworthiness and Efficiency in Explainable AI Systems*, in *2024 IEEE International Conference on Computing (ICOCO)*, 2024, pp. 134–139. doi: 10.1109/ICOCO62848.2024.10928183. Elérhető: https://ieeexplore.ieee.org/document/10928183. [Elérés: 2025. július 10.]
[179] Stoy, Lazarina: *What is an AI winter and is one coming?*, *Search Engine Land*, szept. 2024, Elérhető: https://searchengineland.com/ai-winter-is-coming-446295. [Elérés: 2024. október 1.]
[180] Harguess, Josh & Chris 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, pp. 1–7. doi: 10.1109/AIPR57179.2022.10092230. Elérhető: https://ieeexplore.ieee.org/document/10092230. [Elérés: 2024. október 2.]

[181] Schuchmann, Sebastian: *Analyzing the Prospect of an Approaching AI Winter*, 2019. doi: 10.13140/RG.2.2.10932.91524
[182] Toosi, Amirhosein & mtsai.: *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
[183] Feigenbaum, Edward & Bruce Buchanan & Joshua Lederberg: *On generality and problem solving: A case study using the DENDRAL program*, *Machine Intelligence*, köt. 6, szept. 1970.
[184] Kaplan, Andreas & Michael 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
[185] Nilsson, Nils J.: *The Quest for Artificial Intelligence*, 1. kiad. Cambridge University Press, 2009. doi: 10.1017/CBO9780511819346. Elérhető: https://www.cambridge.org/core/product/identifier/9780511819346/type/book. [Elérés: 2024. október 9.]
[186] Fukushima, Kunihiko: *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, *Biol. Cybernetics*, köt. 36, sz. 4, pp. 193–202, ápr. 1980, doi: 10.1007/BF00344251
[187] Gartner Institute: *Understanding Gartner's Hype Cycles*, *Gartner*, 2018. Elérhető: https://www.gartner.com/en/documents/3887767. [Elérés: 2024. október 8.]
[188] Minsky, M. L. & S. Papert: *Perceptrons, An Essay on Computational Geometry*. Cambridge, MA: MIT Press, 1969.
[189] Haenlein, Michael & Andreas Kaplan: *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*, *California Management Review*, köt. 61, sz. 4, pp. 5–14, aug. 2019, doi: 10.1177/0008125619864925
[190] Lipton, Zachary C. & Jacob 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, pp. Pages 80:45-Pages 80:77, 0 2019, doi: 10.1145/3317287.3328534
[191] Knight, Will: *AI Winter Isn't Coming*, *MIT Technology Review*. Elérhető: https://www.technologyreview.com/2016/12/07/155592/ai-winter-isnt-coming/. [Elérés: 2024. szeptember 30.]
[192] Csepeli, György: *Ember 2.0 - A mesterséges intelligencia gazdasági és társadalmi hatásai*, 2020. kiad. Budapest: Kossuth Kiadó Zrt., Kossuth Kiadó Zrt.
[193] Buck, Anthony R.: *The Way of the Future is now a thing of the past*, *European Academy on Religion and Society*, 2021. március 30. Elérhető: https://europeanacademyofreligionandsociety.com/news/the-way-of-the-future-is-now-a-thing-of-the-past/. [Elérés: 2024. november 9.]
[194] Prisco, Giulio: *About Turing Church*, *Medium*, 2022. október 2. Elérhető: https://turingchurch.net/about-turing-church-ac6ebf2e97b6. [Elérés: 2024. november 11.]
[195] Floridi, Luciano: *AI and Its New Winter: from Myths to Realities*, *Philos. Technol.*, köt. 33, sz. 1, pp. 1–3, márc. 2020, doi: 10.1007/s13347-020-00396-6
[196] Fehér, András Tibor: *A mesterséges intelligencia természetes telei - Az MI technológia lassulását leíró forgatókönyvek és azok védelmi vonatkozásai*, in *A mesterséges intelligencia (MI) hatásainak specifikus vizsgálata*, Dr. Magyar, Sándor & Dr. Kenedli Tamás (szerk.), Budapest: Katonai Nemzetbiztonsági Szolgálat, 2025., pp. 86-125, https://www.knbsz.gov.hu/kiadvanyok#1fd723c5-92d3-4dff-bc47-3edf4cb1b38a [Elérés: 2025. július 12.]

[197] Boncz, Bettina & Roland Szabo: *A mesterséges intelligencia munkaerő-piaci hatásai Hogyan készüljünk fel?*, *Vezetéstudomány / Budapest Management Review*, köt. 53, pp. 68–80, febr. 2022, doi: 10.14267/VEZTUD.2022.02.06
[198] Calegari, Roberta & mtsai.: *Logic-Based Technologies for Intelligent Systems: State of the Art and Perspectives*, *Information*, köt. 11, sz. 3, Art. sz. 3, márc. 2020, doi: 10.3390/info11030167
[199] Wahyono, Aris & mtsai.: *APPLICATION OF CASE BASED REASONING (CBR) METHOD IN DISTRIBUTION TRAFO DAMAGE EXPERT SYSTEM AT PLN UP3 BINJAI*, *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, köt. 2, sz. 1, pp. 1–6, okt. 2022, doi: 10.59934/jaiea.v2i1.111
[200] Aulock, Inge von: *AI Vs. Traditional Methods: A Comprehensive Comparison For Audience Behavior Analysis In 2024*, *PenFriend.AI*, 2024. március 10. Elérhető: https://penfriend.ai/blog/ai-vs-traditional-methods. [Elérés: 2024. július 9.]
[201] Manzoor, Sumaira és mtsai.: *Ontology-Based Knowledge Representation in Robotic Systems: A Survey Oriented toward Applications*, *Applied Sciences*, köt. 11, sz. 10, Art. sz. 10, jan. 2021, doi: 10.3390/app11104324
[202] Ben Said, Aymen & Malek Mouhoub: *A Constraint Satisfaction Problem (CSP) Approach for the Nurse Scheduling Problem*, in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2022, pp. 790–795. doi: 10.1109/SSCI51031.2022.10022250. Elérhető: https://ieeexplore.ieee.org/document/10022250. [Elérés: 2024. július 9.]
[203] Lutkevich, Ben & S. Alexander Gillis: *What are bots and how do they work?*, *WhatIs*. Elérhető: https://www.techtarget.com/whatis/definition/bot-robot. [Elérés: 2024. március 2.]
[204] Thorne, James & mtsai.: *From natural language processing to neural databases*, *Proc. VLDB Endow.*, köt. 14, sz. 6, pp. 1033–1039, febr. 2021, doi: 10.14778/3447689.3447706
[205] Vincent, James: *Forty percent of "AI startups" in Europe don't actually use AI, claims report*, *The Verge*, 2019. március 5. Elérhető: https://www.theverge.com/2019/3/5/18251326/ai-startups-europe-fake-40-percent-mmc-report. [Elérés: 2024. október 16.]
[206] *Mesterséges Intelligencia Elektronikus Almanach*: *26.2. Erős MI: Tudnak-e ténylegesen gondolkodni a gépek?* Elérhető: http://project.mit.bme.hu/mi_almanach/books/aima/ch26s02. [Elérés: 2022. december 16.]
[207] Wang, Pei: *On Defining Artificial Intelligence*, *Journal of Artificial General Intelligence*, köt. 10, sz. 2, pp. 1–37, jan. 2019.
[208] Perrier, Michael Timothy Bennett and Elija: *OpenAI Claims Its New Model Reached Human Level on a Test for 'General Intelligence.' What Does That Mean?*, *Gizmodo*, 2024. december 29. Elérhető: https://gizmodo.com/openai-claims-its-new-model-reached-human-level-on-a-test-for-general-intelligence-what-does-that-mean-2000543834. [Elérés: 2025. január 25.]
[209] Maxwell, Thomas: *Leaked Documents Show OpenAI Has a Very Clear Definition of 'AGI'*, *Gizmodo*, 2024. december 26. Elérhető: https://gizmodo.com/leaked-documents-show-openai-has-a-very-clear-definition-of-agi-2000543339. [Elérés: 2025. január 25.]
[210] Komenczi, Bertalan: *Információelmélet*. Eger: Eszterházy Károly Főiskola, 2011. Elérhető: https://people.inf.elte.hu/molnarba/IREAE _Informacios rendszerek elmeleti alapjai EA. (MSc Ir)/Informaciorendszerek elmeleti alapjai/Books/Magyar Forrasok/0005_20_informacioelmelet_pdf.pdf [Elérés: 2024. január 10.]

[211] Shannon, C. E.: *A Mathematical Theory of Communication*, *Bell System Technical Journal*, köt. 27, sz. 3, pp. 379–423, júl. 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x
[212] Otten, Klaus & Anthony Debons: *Towards a metascience of information: Informatology*, *Journal of the American Society for Information Science*, köt. 21, sz. 1, pp. 89–94, 1970, doi: 10.1002/asi.4630210115
[213] Borko, Herald: *Information science: What is it?*, *American Documentation*, köt. 19, sz. 1, pp. 3–5, 1968, doi: 10.1002/asi.5090190103
[214] Linares Columbié, Radamés: *Harold Borko y la Ciencia de la Información*, *Revista Cubana de Información en Ciencias de la Salud*, köt. 27, sz. 3, pp. 410–419, szept. 2016.
[215] Murányi Péter: *Az információtudomány eredete, fejlődése és kapcsolatai*, *Tudományos és Műszaki Tájékoztatás*, köt. 41, sz. 7–8, Art. sz. 7–8, 1994.
[216] Könyvtári Figyelő Szerkesztősége: *Az információtudomány létrejötte és Horváth Tibor információtudományi eszméinek gyökerei1 | Könyvtári Figyelő*. Elérhető: http://ki2.oszk.hu/kf/2017/04/az-informaciotudomany-letrejotte-es-horvath-tibor-informaciotudomanyi-eszmeinek-gyokerei1/. [Elérés: 2023. december 26.]
[217] Khanye, Nqobulwazi: *Information science: origin, evolution and relations*, *Conceptions of library and information science: …*, jan. 1992, Elérhető: https://www.academia.edu/925298/Information_science_origin_evolution_and_relations. [Elérés: 2023. december 27.]
[218] Stock, Wolfgang G. & Mechthild Stock: *Handbook of information science*. Berlin Boston: De Gruyter Saur, 2013.
[219] Boros, János: *A kognitív tudomány esélyei*, *Magyar Tudomány*, sz. 2004/11, pp. 1269, 2004.
[220] Sisu, Diana: *School of Informatics: in memoriam Hugh Christopher Longuet-Higgins*. Elérhető: https://www.inf.ed.ac.uk/events/christopherlh.html. [Elérés: 2023. december 28.]
[221] Flowers, B. H.: *Lighthill Report: Artificial Intelligence: a paper symposium*, UKRI Science and Technology Facilities Council, 1973. Elérhető: https://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm. [Elérés: 2023. december 28.]
[222] Kodaj, Dániel: *A kognitív tudomány mitikus teleológiája*, *Magyar Tudomány*, sz. 10, pp. 1250, 2005.
[223] Pléh, Csaba: *A tudomány jövője: A kognitív tudomány példája a tudományok tagolódásáról és diverzifikálódásáról*, *Magyar Tudomány*, köt. 168, sz. 9, Art. sz. 9, 2007.
[224] Kovács, Győző & Endre Tóth: *Az informatika története*. Budapest: NJSZT Informatikatörténeti Fórum, 1991. Elérhető: https://itf.njszt.hu/objektum/az-informatika-tortenete. [Elérés: 2023. december 28.]
[225] Biener, Klaus: *Karl Steinbuch – Informatiker der ersten Stunde*, *RZ-Mitteilungen*, sz. 15, pp. 53–54, 1997.
[226] Steinbuch, Karl: *Informatik: Automatische Informationsverarbeitung*, *SEG-Nachrichten (Technische Mitteilungen der Standard Elektrik Gruppe) – Firmenzeitschrift*, sz. 4, pp. 171, 1957.
[227] Daum, Silke (ITIV) (inaktiv): *KIT - ITIV - Institut - Geschichte - Geschichte des ITIV - Karl Steinbuch - Gründer des ITIV*, 2021. december 14. Elérhető: https://www.itiv.kit.edu/4787.php. [Elérés: 2023. december 26.]

[228] „L’équipe Ça m’intéresse” csapat: *Quelle est l’origine du mot informatique ?*, *Ça m’intéresse*, 2020. április 6. Elérhető: https://www.caminteresse.fr/societe/quelle-est-lorigine-du-mot-informatique-11137313/. [Elérés: 2023. december 26.]
[229] Masic, Izet: *The Most Influential Scientists in the Development of Biomedical Informatics: Philippe Dreyfus (1925-2018)*, köt. 10, pp. 137–137, szept. 2022, doi: 10.5455/ijbh.2022.2.137-137
[230] IT History Society: *Dr. Walter F. Bauer*, 2015. december 21. Elérhető: https://www.ithistory.org/honor-roll/dr-walter-f-bauer. [Elérés: 2023. december 28.]
[231] Shaoyi, HE & István (ford) Papp: *Az informatika fogalma*, *TMT*, köt. 21, sz. 2, pp. 117–122, 2003.
[232] Fichman, Pnina & Rosenbaum Howard: *Social Informatics: Past, Present and Future*. Newcastle: Cambridge Scholars Publishing, 2014. Elérhető: https://www.cambridgescholars.com/product/978-1-4438-5576-1. [Elérés: 2023. december 26.]
[233] Collins English Dictionary: *Informatics definition and meaning* |, 2023. december 28. Elérhető: https://www.collinsdictionary.com/dictionary/english/informatics. [Elérés: 2023. december 28.]
[234] Cambridge Dictionary: *Informatics*, 2023. december 27. Elérhető: https://dictionary.cambridge.org/dictionary/english/informatics. [Elérés: 2023. december 28.]
[235] Munk, Sándor: *A katonai informatika alapelvei a magyar honvédségben i. (alapok)*, *Hadmérnök*, köt. IV, sz. 3, 2009, Elérhető: http://www.hadmernok.hu/2009_3_munk1.php. [Elérés: 2023. december 28.]
[236] The University of Edinburgh: *Looking back to the future of Edinburgh’s AI legacy*, , 2023. május 5. Elérhető: https://www.ed.ac.uk/bayes/media-centre/news/ai-news/looking-back-to-the-future-of-edinburghs-ai-legacy. [Elérés: 2023. december 28.]
[237] The University of Edinburgh: *What is Informatics?*, *The University of Edinburgh*, 2023. január 12. Elérhető: https://www.ed.ac.uk/informatics/about/what-is-informatics. [Elérés: 2023. december 28.]
[238] Collins, Richard: *Back to the future: Digital television and convergence in the United Kingdom*, *Telecommunications Policy*, köt. 22, sz. 4, pp. 383–396, máj. 1998, doi: 10.1016/S0308-5961(98)00022-6
[239] Khan, Hanif: *Types of AI | Different Types of Artificial Intelligence Systems*, *Fossguru*, köt. 9, pp. 50, szept. 2021.
[240] 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. Elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX%3A52021PC0206. [Elérés: 2024. február 4.]
[241] Regona, Massimo & mtsai.: *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
[242] Browne, Ryan: *Italy became the first Western country to ban ChatGPT. Here’s what other countries are doing*, *CNBC*, 2023. április 4. Elérhető: https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html. [Elérés: 2025. február 13.]
[243] Resperger, István: *A válságkezelés és a hibrid hadviselés*. Budapest: Dialóg Campus, 2018. Elérhető: https://tudasportal.uni-nke.hu/xmlui/handle/20.500.12944/12438. [Elérés: 2023. február 16.]

[244] Kiss, Álmos Péter: *A hibrid hadviselés természetrajza*, *Honvédségi Szemle*, köt. XI, sz. 4, Art. sz. 4., 2019.
[245] Porkoláb, Imre: *Hibrid hadviselés: új hadviselési forma, vagy régi ismerős?*, *Hadtudomány*, köt. 25, sz. 3–4, Art. sz. 3–4, 2015.
[246] Albergotti, Reed & Drew Harwell: *Apple and Google are building a virus-tracking system. Health officials say it will be practically useless.*, *Washington Post*, 2020. május 15. Elérhető: https://www.washingtonpost.com/technology/2020/05/15/app-apple-google-virus/. [Elérés: 2023. január 28.]
[247] Mantellassi, Federico: *Digital Authoritarianism: How Digital Technologies Can Empower Authoritarianism and Weaken Democracy*, *Geneva Centre for Security Policy*, 2023, Elérhető: https://www.gcsp.ch/publications/digital-authoritarianism-how-digital-technologies-can-empower-authoritarianism-and. [Elérés: 2023. március 2.]
[248] Krajnc, Zoltán (szerk.): A *hadművelet célja*, *Hadtudományi lexikon*. Dialóg Campus, Budapest, 2019. Elérhető: https://rendeszet.uni-nke.hu/publikaciok/hadtudomanyi-lexikon. [Elérés: 2022. december 23.]
[249] Kovács, László: *Offenzív kiberműveletek 1.: Az offenzív kiberműveletek természete*, *Hadmérnök*, köt. 16, sz. 2, Art. sz. 2, aug. 2021, doi: 10.32567/hm.2021.2.13
[250] Boda, 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
[251] *Háború a szürke zónában - Kiss Álmos Péterrel, az MH Transzformációs Parancsnokság Honvéd Tudományos Kutatóhely vezető kutatójával a hibrid hadviselésről beszélgettünk.*, 2021. december 6. Elérhető: https://honvedelem.hu/hirek/haboru-a-szurke-zonaban.html. [Elérés: 2022. november 19.]
[252] Haig, Zsolt: *Kibertéri kognitív befolyásolás az információs műveletekben*, *Hadtudományi Szemle*, köt. 15, sz. 2, Art. sz. 2, okt. 2022, doi: 10.32563/hsz.2022.2.7
[253] Somodi, Zoltán & Kiss Álmos Péter: *A hibrid hadviselés fogalmának értelmezése a nemzetközi szakirodalomban*, *Honvédségi Szemle – Hungarian Defence Review*, köt. 147, sz. 6, Art. sz. 6, 2019.
[254] Moravec, Hans: *Bodies, Robots, Minds*. Robotics Institute, 1995. Elérhető: https://frc.ri.cmu.edu/~hpm/project.archive/general.articles/1995/Kunstforum.html. [Elérés: 2019. június 1.]
[255] Taksás, Balázs: *Újjászülető hadiipar*, *HADITECHNIKA*, köt. 55, sz. 1, Art. sz. 1, 2021.
[256] Bihaly Barbara: *A mesterséges intelligencia felhasználása az információs és kibertérműveletekben – az orosz minta*, *Hadmérnök*, köt. 17, sz. 3, Art. sz. 3, nov. 2022, doi: 10.32567/hm.2022.3.7
[257] Dawson, Andrew & Martin Innes: *How Russia's internet research agency built its disinformation campaign*, *Political Quarterly*, köt. 90, sz. 2, Art. sz. 2, jún. 2019.
[258] Department of Defence Test Method Standard: *MIL-STD-810: Environmental Engineering Considerations and Laboratory Tests*. in US Military. DdD - US Military, 2000. Elérhető: https://www.atec.army.mil/publications/mil-std-810g/mil-std-810g.pdf [Elérés: 2023. január 22.]
[259] Stiglitz, Joseph E.: *How the US Could Lose the New Cold War*, *Project Syndicate*, 2022. június 17. Elérhető: https://www.project-syndicate.org/commentary/us-squandering-soft-power-appeal-in-cold-war-with-china-by-joseph-e-stiglitz-2022-06. [Elérés: 2022. november 19.]
[260] Csirmaz, Evelin: *LAWS - a gyilkos robotok és a nemzetközi jog*, *arsboni*, 2018. november 8. Elérhető: https://arsboni.hu/laws-a-gyilkos-robotok-es-a-nemzetkozi-jog/. [Elérés: 2022. november 19.]

[261] Porkoláb, Imre & Négyesi Imre: *A mesterséges intelligencia alkalmazási lehetőségeinek kutatása a haderőben*, *Honvédségi Szemle*, köt. 147, sz. 5., Art. sz. 5., 2019.
[262] Fehér, András Tibor & Négyesi Imre: *Mesterségesintelligencia-alapú kibertértámadási modellek*, *MKK*, köt. 31, sz. 3, pp. 73–87, 2021, doi: 10.32562/mkk.2021.3.5
[263] Rijmenam, Mark van: *What is Generative AI, and How Will It Disrupt Society?*, *The Digital Speaker: Futurist, Keynote Speaker, Author*, 2022. november 11. Elérhető: https://www.thedigitalspeaker.com/what-is-generative-ai-how-disrupt-society/. [Elérés: 2022. december 20.]
[264] Ata, Nabil Abu el: *What is Generative Intelligence?*, *URM GROUP*, 2018. január 9. Elérhető: https://urmgrp.com/what-is-generative-intelligence/. [Elérés: 2022. december 20.]
[265] Fehér, András Tibor & Négyesi Imre: *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
[266] Stoecklin, Marc & Ph.Jang Jiyong & Dhilung Kirat: *DeepLocker: How AI Can Power a Stealthy New Breed of Malware*. 2018. Elérhető: https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/ [Elérés: 2022. április 2.]
[267] Kirat, Dhilung & G. Vigna & C. Kruegel: *BareBox: Efficient Malware Analysis on Bare-Metal*, in *Twenty-Seventh Annual Computer Security Applications Conference, ACSAC (old*, Orlando, USA and Letlts dtuma: 403–412) and ACM and 2019, 2011.
[268] Muha, Lajos & Csaba Krasznay: *Az elektronikus információs rendszerek biztonságának menedzselése*. Budapest: Nemzeti Közszolgálati Egyetem, 2018.
[269] Fabók, Bálint: *Magyar céggel csaltak el tízmilliókat az első európai hanghamisításos átverésnél*, *G7.hu*, 2019, Elérhető: https://g7.hu/tech/20190903/magyar-ceggel-csaltak-el-tizmilliokat-az-elso-europai-hanghamisitasos-atveresnel/ [Elérés: 2022. április 20.]
[270] Kovács, László: *A kibertér védelme*. in Pro patria ad mortem. Budapest: Dialóg Campus, 2018.
[271] Da Silva, Wilson: *War of the Bots: Artificial Intelligence in Cyber Warfare*, *Dialogue & Discourse*, 2019, Elérhető: https://medium.com/discourse/spy-vs-spy-cyber-warfare-gets-automated-aba60ece738c [Elérés: 2022. április 21.]
[272] Huang, W. Ronny & mtsai.: *MetaPoison: Practical General-purpose Clean-label Data Poisoning*, *Advances in Neural Information Processing Systems*, köt. 33, pp. 12080–12091, 2020.
[273] Durbin, Steve: *How Criminals Use Artificial Intelligence To Fuel Cyber Attacks*, *Forbes*, 2020, Elérhető: https://www.forbes.com/sites/forbesbusinesscouncil/2020/10/13/how-criminals-use-artificial-intelligence-to-fuel-cyber-attacks/?sh=5e4c8ac85012 [Elérés: 2022. április 19.]
[274] Dixon, William & Jamil Farshchi: *AI is the latest weapon cybercriminals are exploiting*. 2019. Elérhető: https://www.weforum.org/agenda/2019/09/4-ways-ai-is-changing-cybersecurity-both-in-attack-and-defense/ [Elérés: 2022. április 10.]
[275] Cem, Dilmegani: *24 Affective Computing (Emotion AI) Applications / Use Cases*. 2020. Elérhető: https://research.aimultiple.com/affective-computing-applications/ [Elérés: 2023. április 24.]
[276] Diederich, Stephan & mtsai.: *Emulating Empathetic Behavior in Online Service Encounters with Sentiment-Adaptive Responses: Insights from an Experiment with a Conversational Agent*, *ICIS 2019 Proceedings*, 2019, Elérhető:

https://aisel.aisnet.org/icis2019/smart_service_science/smart_service_science/2 [Elérés: 2022. április 8.]
[277] Defense Advanced Research Projects Agency: *Department of DefenseFiscal Year (FY) 2020 Budget Estimates*. Washington: Defense Advanced Research Projects Agency, 2019.
[278] Négyesi, Imre: *A mesterséges intelligencia és az etika*, *Hadtudomány*, köt. 30, pp. 103-113., 2020, doi: 10.17047/HADTUD.2020.30.1.103
[279] Europe Parlament & Science Communication Unit 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. Elérhető: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)634452. [Elérés: 2022. december 26.]
[280] Valpiani, James: *ACE: Air Combat Evolution*, 2024. Elérhető: https://www.darpa.mil/research/programs/air-combat-evolution. [Elérés: 2025. július 29.]
[281] Wang, Guan és mtsai.: *Hierarchical Reasoning Model*. arXiv, 2025. augusztus 4. doi: 10.48550/arXiv.2506.21734. Elérhető: http://arxiv.org/abs/2506.21734. [Elérés: 2025. augusztus 6.]
[282] Chi, Zhiming és mtsai.: *Patch Synthesis for Property Repair of Deep Neural Networks*. arXiv, 2024. február 1. doi: 10.48550/arXiv.2404.01642. Elérhető: http://arxiv.org/abs/2404.01642. [Elérés: 2024. augusztus 7.]
[283] Kim, Hoon-Hee & Jaeseung Jeong: *Recovery of computation capability for neural networks from damaged states using self-organized criticality*, *BMC Neurosci*, köt. 11, sz. Suppl 1, pp. P30, júl. 2010, doi: 10.1186/1471-2202-11-S1-P30
[284] 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. 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 [Elérés: 2023. június 20.]
[285] Miniszterelnöki Kabinetiroda: *Nemzeti Digitalizációs Stratégia (2022-2030)*, Budapest, 2022. Elérhető: https://kormany.hu/dokumentumtar/nemzeti-digitalizacios-strategia-2022-2030. [Elérés: 2023. március 2.]
[286] Zaman, Rokon: *Nokia's Collapse: Lessons in Reinvention, Disruption and Adaptation*, *THE WAVES*, 2024. október 13. Elérhető: https://www.the-waves.org/2024/10/13/nokias-collapse-lessons-in-reinvention-disruption-and-adaptation/. [Elérés: 2025. február 10.]
[287] Sherman, Justin: *India's Digital Path: Leaning Democratic or Authoritarian?*, *Just Security*, 2019. február 4. Elérhető: https://www.justsecurity.org/62464/indias-digital-path-leaning-democratic-authoritarian/. [Elérés: 2022. november 19.]
[288] Yayboke, Erol & Samuel Brannen: *Promote and Build: A Strategic Approach to Digital Authoritarianism*, Center for Strategic and International Studies (CSIS), Washington, 2020. Elérhető: https://www.csis.org/analysis/promote-and-build-strategic-approach-digital-authoritarianism. [Elérés: 2022. december 24.]
[289] Morgus, Robert: *The Spread of Russia's Digital Authoritarianism*, in *Artificial Intelligence, China, Russia, and the Global Order*, Air University Press, 2019, pp. 89–97. Elérhető: https://www.jstor.org/stable/resrep19585.17. [Elérés: 2023. március 17.]
[290] Meserole, Chris & Alina Polyakova: *Exporting digital authoritarianism -The Russian and Chinese models*, *Brookings*, 2019. augusztus 26. Elérhető:

https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf. [Elérés: 2022. november 18.]
[291] Gigova, Radina: *Who Vladimir Putin thinks will rule the world*, *CNN*, szept. 2017, Elérhető: https://www.cnn.com/2017/09/01/world/putin-artificial-intelligence-will-rule-world/index.html. [Elérés: 2022. november 22.]
[292] Emelin, Konstantin: *Artificial Intelligence is Under Control – Комиссия Российской Федерации по делам ЮНЕСКО*, *UNESCO*, 2020. Elérhető: http://unesco.ru/en/news/ai-committee/. [Elérés: 2023. március 29.]
[293] Soldatov, Andrei & Irina Borogan: *5 Russian-Made Surveillance Technologies Used in the West*, *Wired*. Elérhető: https://www.wired.com/2013/05/russian-surveillance-technologies/. [Elérés: 2023. március 17.]
[294] Zakharov, Andrej: *Злость, страх и силуэты. Мэрия Москвы раскрыла, какие алгоритмы распознают людей по лицам*, *BBC News Русская служба*, 2022. augusztus 25. Elérhető: https://www.bbc.com/russian/features-62658404. [Elérés: 2022. november 27.]
[295] Statista elemző oldal: *CCTV camera density by major city Russia 2022*, *Statista*, 2022. 0. Elérhető: https://www.statista.com/statistics/1156026/surveillance-cameras-density-moscow-st-petersburg/. [Elérés: 2022. november 27.]
[296] Bendett, Samuel: *Russia's Artificial Intelligence Boom May Not Survive the War*. Elérhető: https://www.defenseone.com/ideas/2022/04/russias-artificial-intelligence-boom-may-not-survive-war/365743/. [Elérés: 2022. november 19.]
[297] Bendett, Samuel: *Russia's AI Quest is State-Driven — Even More than China's. Can It Work?*, *Defense One*. Elérhető: https://www.defenseone.com/ideas/2019/11/russias-ai-quest-state-driven-even-more-chinas-can-it-work/161519/. [Elérés: 2022. november 19.]
[298] Palavenis, Donatas: *The Use of Emerging Disruptive Technologies by the Russian Armed Forces in the Ukrainian W*, *Air Land Sea Space Application (ALSSA) Center*, 2022. Elérhető: https://www.alsa.mil/Portals/9/Documents/articles/221001_ALSA_Article_Donatas_Palavenis.pdf. [Elérés: 2023. március 29.]
[299] RosBusinessConsulting híroldal: *В России протестировали работу Рунета при отключении от глобальной Сети*, *РБК*, 2021. Elérhető: https://www.rbc.ru/technology_and_media/21/07/2021/60f8134c9a79476f5de1d739. [Elérés: 2022. november 26.]
[300] Pravda.ru: *Российский рунет выдержал первые испытания*, *pravda.ru*, 2019. december 23. Elérhető: https://www.pravda.ru/news/politics/1461663-runet/. [Elérés: 2022. november 26.]
[301] Criddle, Cristina: *Russia slows down Twitter over „banned content"*, *BBC News*, 2021. március 10. Elérhető: https://www.bbc.com/news/world-europe-56344304. [Elérés: 2022. november 19.]
[302] Rose, Janus: *Russia's Digital Iron Curtain Is Starting To Take Shape*, *Vice*, 2022. március 17. Elérhető: https://www.vice.com/en/article/5dg4kb/russias-digital-iron-curtain-is-starting-to-take-shape. [Elérés: 2022. november 26.]
[303] Sinkkonen, Elina & Jussi Lassila: *Digital Authoritarianism and Technological Cooperation in Sino-Russian Relations: Common Goals and Diverging Standpoints*, in Kirchberger, S. & S. Sinjen, & N. Wörmer, (szerk.): *Russia-China Relations: Emerging Alliance or Eternal Rivals?*, Global Power Shift. Cham: Springer International Publishing, 2022, pp. 165–184. doi: 10.1007/978-3-030-97012-3_9. Elérhető: https://doi.org/10.1007/978-3-030-97012-3_9. [Elérés: 2022. november 18.]

[304] Lalljee, Jason: *Russia's economy already lost $860 million this year because the government keeps shutting down the internet*, *Business Insider*, 2022. március 19. Elérhető: https://www.businessinsider.com/russia-internet-censorship-cost-economy-putin-ukraine-sanctions-twitter-2022-3. [Elérés: 2023. március 22.]
[305] TASS hírügynökség: *Russia won't be left without Internet, guarantees Lavrov*, *TASS*. Elérhető: https://tass.com/politics/1452143. [Elérés: 2022. november 26.]
[306] Freedom House civil szervezet: *Russia: Freedom on the Net 2022 Country Report*, *Freedom House*. Elérhető: https://freedomhouse.org/country/russia/freedom-net/2022. [Elérés: 2022. november 26.]
[307] Toulas, Bill: *Russia creates its own TLS certificate authority to bypass sanctions*, *BleepingComputer*, 2022. Elérhető: https://www.bleepingcomputer.com/news/security/russia-creates-its-own-tls-certificate-authority-to-bypass-sanctions/. [Elérés: 2022. november 26.]
[308] Claburn, Thomas: *Russia bans foreign software purchases for infrastructure*, *The Register*, 2022. Elérhető: https://www.theregister.com/2022/04/01/russia_bans_foreign_software/. [Elérés: 2023. március 17.]
[309] Hughes, Matthew: *Microsoft stops serving Windows downloads to Russian users*, *KnowTechie*, 2022. június 21. Elérhető: https://knowtechie.com/microsoft-stops-serving-windows-downloads-to-russian-users/. [Elérés: 2023. március 17.]
[310] Howells, Laura & Laura A. Henry: *Digital Authoritarianism at War: Controlling Russia's Information Space*, *NYU Jordan Center*, 2022. április 26. Elérhető: https://jordanrussiacenter.org/news/digital-authoritarianism-at-war-controlling-russias-information-space/. [Elérés: 2022. november 27.]
[311] Harding, Luke: *Twitter launches privacy-protected site on dark web to bypass Russia's block*, *The Guardian*, 2022. március 10. Elérhető: https://www.theguardian.com/technology/2022/mar/09/twitter-tor-version-russia-block. [Elérés: 2022. november 18.]
[312] McGiffin, Joseph M.: *Mission (Command) Complete: Implications of JADC2*, *National Defense University Press*, sz. 2024, pp. July, 2024.