CSÁKI CSABA

A MESTERSÉGES INTELLIGENCIA ELTERJEDÉSÉBŐL ADÓDÓ KOCKÁZATOK SZISZTEMATIKUS VIZSGÁLATA

Bevezetés

Jelen kutatás a mesterséges intelligencia (MI) úgynevezett keskeny (*narrow* – magyarul „gyenge", „szűk" vagy „speciális") formájára fókuszál, ahol a mesterséges keskeny intelligencia (MKI) kifejezés a speciális céllal készült, adatvezérelt, modellalapú intelligens rendszereket jelenti, amelyek jellemzően „*data science*" megközelítésben, gépi tanulási módszerekre építve készülnek. Ezeket a megoldásokat szervezeti környezetben tipikusan döntéshozó és döntéstámogató, illetve kommunikációs feladatokra használják. A téma aktualitását az adja, hogy a gépi tanulási rendszereket egyre szélesebb körben és egyre nagyobb számban alkalmaznak olyan, akár nyitott szituációkban is, ahol e rendszerek autonóm módon viselkedhetnek.

Ugyanakkor a mesterséges intelligenciát (MI) érintő közbeszédet, de a szakmai vitákat és a terület tudományos vizsgálatát is jelenleg néhány téma erős túlsúlya jellemzi. A viták középpontjában a jelenséghez kapcsolódó etikai kérdések állnak. Jogi oldalról a szabályozás problematikájának boncolgatása figyelhető meg, amely elsősorban a személyes adatokhoz (*privacy*) kapcsolódó jogok és kötelességek viszonyrendszerére koncentrál, ahol inkább problémák és kihívások dzsungeleként tekintenek a területre, míg a megoldásokra vonatkozó javaslatok viszonylag szűk körben mozognak. Szakmai oldalról még mindig a hurráoptimista innovációs lendület a jellemző, a felvetődő etikai-erkölcsi kihívásokat gyakran felszínes látszatmegoldásokkal (iránymutatások, etikai kódexek, etikai bizottságok és laza szerveződések súlytalan egyvelegével) igyekeznek mederben tartani. A tudományos világ alkalmanként szélsőségekbe esve és egzisztenciális veszélyektől aggódva, máskor rácsodálkozással szemléli a lehetséges jövőt, a megoldások tekintetében pedig szintén naiv optimizmussal hisz a felmerülő problémák kezelhetőségében, és rendszeresen megmarad a technológiai részletekben rejlő problémák aprólékos vizsgálatánál.

Az etikai szabályozási kérdések mellett több dimenzióban (a szabályozás, a közbeszéd és a szakma vitatereiben) rendszeresen felmerül a munkára, munkaerőre gyakorolt hatás,[1] amelynek vizsgálatából a nagy tanácsadó cégek is kiveszik a részüket. A lehetséges negatív következményeket vizsgáló mélyebb elemzések is sokszor csak bizonyos területekre koncentráltak, és jellemzően általános (és negatív) társadalmi hatásokra, a munkaerőpiac egyes szegmenseit érintő változásokra, vagy

[1] MIRBABAIE, Milad – BRÜNKER, Felix – MÖLLMANN, Nicholas R. J. (Frick) – STIEGLITZ, Stefan: The Rise of Artificial Intelligence – Understanding the AI Identity Threat at the Workplace. Electronic Markets, Volume 32, Issue 1, 2022. pp. 73–99.

iparághoz kötődő veszélyekre figyelmeztettek. A nagyon erős mesterséges intelligencia esetében is hasonló a helyzet. Ebben a közegben azonban hiányzik a kockázatok átfogó, szisztematikus elemzése. A jelenleg domináns keskeny MI esetében nem ismert olyan kutatás, amely egy általános elméleti keretben próbálná meg elhelyezni és vizsgálni a kapcsolódó kockázatokat. De nem készült még a megelőzési lehetőségek szisztematikus, egységes, és integrált vizsgálatát célul kitűző empirikus (tehát nem csak elméleti fejtegetésekre épülő) kutatás sem – vagy az eredmények védettek.

E tanulmány bemutat és empirikus eredményekre építve validál egy olyan egységes MKI kockázatelemzési keretrendszert, amely akár adott területek mentén, akár integráltan képes releváns kockázatok beazonosítására és menedzselésére. A mesterségesintelligencia-ökoszisztéma kidolgozott modellje segítségével továbbá beazonosíthatók azok a pontok, ahol a védekezés vagy megelőzés jó eséllyel megvalósítható, illetve a rendelkezésre álló megelőzési lehetőségek közül a legelőnyösebb megoldásra vagy azok kombinációjára vonatkozó javaslatok kidolgozhatók. A kutatás fontos eleme a többi empirikus tématerület eredményeivel történő összehangolás, az ott felmerülő specifikus kockázatok taxonómiai vizsgálata, illetve a megelőzési lehetőségek különböző dimenziók (technikai, regulációs, szervezeti, oktatási, társadalmi stb.) mentén történő értelmezése. Az integrációt az ökoszisztéma-alapú megközelítés biztosítja. A kutatási jelentés három fő részből áll:

1) általános kockázatelemzési elméletek és szakmai gyakorlatok, illetve specifikus MI-kockázatok áttekintése tudományos, szakmai és internetes források segítségével;
2) a keskeny MI mint információrendszerhez kapcsolódó munkafolyamat vizsgálatán keresztül egy egységes, ökoszisztéma-alapú kockázatelemzési keretrendszer kidolgozása;
3) az MI-hez kapcsolódható kockázatok értelmezése, taxonómiai elemzése, és ezzel párhuzamosan a megelőzés és a kezelés jelenlegi lehetőségeinek összefoglalása.

Kockázat és biztonság az általános és az IT-szakirodalomban

A kockázat legegyszerűbb definíciója szerint egy nem kívánatos állapot és az annak előfordulásához kapcsolódó valószínűség. Ugyanakkor gyakorlati szempontból a kockázatnak többféle értelmezése lehet. A fenti veszélyhelyzet-értelmezés mellett egyrészt vizsgálható a kockázat mint a biztonság hiánya, másrészt a gyakorlatban megfigyelhető a kockázatok szakmaspecifikus vagy iparágra szabott elemzése, menedzselése.

A biztonság alapdefiníciója abból a feltevésből indul ki, hogy az emberek biztonságban, megbízható (társadalmi-gazdasági-emberi) környezetben akarnak élni.[2] Akkor érzik magukat biztonságban, ha bizonyos negatív hatású események valószínűsége alacsony, vagy legalábbis egy számukra elfogadható szint alatt marad.

[2] GASPER, Des – GÓMEZ, Oscar A.: Human security thinking in practice: 'personal security', 'citizen security' and comprehensive mappings. Contemporary Politics, Volume 21, Issue 1, 2015. pp. 100–116.

Ilyen esemény tipikusan, ami a) veszélyezteti a mindennapjaikat (munkájukat, tulajdonukat, fizikai létezésüket); b) megsérti a magánszférájukat; vagy c) félelmet kelt. Hétköznapi értelemben a biztonságérzet azt is magában foglalja, hogy segítségre számíthatnak, ha baj van (mint pl. baleset, betegség, természeti katasztrófa). Habár személyes vagy társadalmi érzetről van szó, a biztonság mégis egy (objektíven) kívánatos állapot, amelyet valamennyien szeretnénk elérni és megtartani.[3] Természetesen a kockázathoz hasonlóan a biztonságnak is lehetnek eltérő értelmezései és dimenziói – a magánvonatkozás mellett társadalmi, szervezeti vagy például IT-biztonságról is beszélhetünk.

Tehát a biztonság eléréséhez két összekapcsolódó fogalom figyelembevétele alapvető: 1) magának a biztonságnak mint kívánatos állapotnak a leírása; és 2) a kockázatnak mint a kívánatos állapot megszűnéséhez kapcsolódó valószínűségnek a fogalma. Itt fontos megjegyezni, hogy ebből a kettős megközelítésből nézve a biztonság nem egyszerűen kockázatkerülést jelent, hanem a kockázatok menedzselését, ami feltételezi azok lehetőségekhez mért minél részletesebb feltérképezését. Azaz itt visszatértünk az eredeti kiinduló definícióhoz.

Általános ipari technológiák esetén a kvantitatív, valószínűségeket és lehetséges veszteségeket számszerűsíteni igyekvő megközelítések mellett egyre jobban megjelennek a kvalitatív és az emberi tényezőket is figyelembe vevő kockázatelemzési keretek.[4] Ezek a szociotechnikai megközelítések vizsgálják a technológia helyét és szerepét a rendszeren belül, és megpróbálják feltárni, hogy kik és milyen módon kerülnek kapcsolatba az adott technológiával. Ebben a viszonyrendszerben kockázati mutatókat állítanak fel, amelyeket aztán a kockázat kezelésére igyekeznek felhasználni – a technológia működtetésének közegében. E mutatókat rendszeresen mérni és felügyelni kell, miközben hagyományos ellenőrzési mechanizmusokat kapcsolnak hozzájuk. Az ipari szakirodalom számos főbb kockázatbefolyásoló keretrendszert ismer, amelyek többsége a szociotechnikai rendszerben végrehajtható tevékenységek hatását igyekszik felmérni a kapcsolódó és felmért kockázatok szintjére. Mind a befolyásoló tényezők, mind a kockázatok számos dimenzió mentén vizsgálhatók: szervezeti, működési, személyes, technikai, feladat és környezeti elemek jelenhetnek meg egy-egy keretrendszer vizsgálódási körében. E keretek problémája, hogy a szociotechnikai megközelítés ellenére hajlamosak mégis inkább kvantitatív indikátorokra fókuszálni, de ezeket a mutatókat nem kötik össze rendszerszinten a kockázatok felmérésével, csak a biztonságra gyakorolt hatásokkal. Ugyanakkor erősségük az incidensek kiváltó okainak egyértelmű feltárása és az emberi tényezők kezelésnek előtérbe helyezése.[5]

---

[3] ISMAGILOVA, Elvira – HUGHES, Laurie – RANA, Nripendra P. – DWIVEDI, Yogesh K.: Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. Information Systems Frontiers, Volume 24, Issue 2, 2022. pp. 393–414.

[4] PENCE, Justin – MOHAGHEGH, Zahra – OSTROFF, Cheri – LEE, Ernie – YILMAZ, Fatima – GRANTOM, Rick – JOHNSON, David: Toward monitoring organizational safety indicators by integrating probabilistic risk assessment, socio-technical systems theory, and big data analytics. 12th Int. Probabilistic Safety Assessment and Management Conference, 2014. pp. 237–251.

[5] ØIEN, Knut: A framework for the establishment of organizational risk indicators. Reliability Engineering & System Safety, Volume 74, Issue 2, 2001. pp. 147–167.

Az integrált megközelítés szerint[6] a kockázatok felmérése (egy adott szervezeti közeget vizsgálva) három lépést feltételez: a kockázat felismerése, a kockázat elemzése, a kockázat értékelése. Ezek mindegyike további résztevékenységeket fed le az alábbiak szerint. A *felismerést* segíti a kockázati okok és források „felfedezése", veszélyek és gyengeségek keresése, a lehetőségek és a képességek felmérése, illetve a rendelkezésre álló erőforrások természetének és értékének ismerete. Az *elemzés* célja a kockázati szintek megállapítása, amelyhez szükséges a hibák bekövetkezési valószínűségéhez és következményeinek megértéséhez szükséges információk összegyűjtése és felhasználása, illetve a következmények természetének és terjedelmének megbecsülése a rendelkezésre álló információk alapján. Az első két lépés eredményét felhasználva kerülhet sor a kockázatok *értékelésére* úgy, hogy a döntéshozók számára a kvantitatív rálátás mellett meg kell határozni a kockázatok elfogadható szintjét is.

Specifikusan tekintve az informatika vagy információtechnológia (IT) területét, egyrészt a COBIT 5.0[7] kockázatalapú IT-auditra vonatkozó részei szolgálhatnak tájékozódási pontként, másrészt pedig az ISACA CISA Review Manual[8] kockázatmenedzselési folyamatra vonatkozó javaslata ajánl egy keretrendszert az IT-kockázatok kezelésére (lásd például a Manual 1.3 ábráját). Az előbbi a szervezeti IT-alkalmazások folyamatos felügyeletére és rendszeres, szisztematikus auditjára helyezi a hangsúlyt, az utóbbi lényegében egy periódikusan végrehajtandó kockázatértékelési folyamatot ajánl, amelynek lépései: üzleti célok azonosítása, az üzleti célokat támogató információvagyon azonosítása, kockázatok felmérése (a veszély-sérülékenység-valószínűség-hatás logikája mentén), a kockázatokat (azok valószínűségét és hatását is) mérséklő kontrollok hozzárendelése az azonosított kockázatokhoz (egyrészt a meglévő kontrollok kockázatokhoz rendelése, másrészt a nem vagy nem eléggé kezelt kockázatokhoz új kontrollok kialítása).

Kockázat és biztonság az MI szakirodalomában

Az általános kockázatvizsgálati megközelítésekhez hasonlóan az MI-hez kapcsolható kockázatok is értelmezhetők az egyén, a szervezet és a társadalom szintjén, illetve a kockázatok osztályozása lehetséges az érintett (társadalmi-emberi) területek szerint is. Emellett mind a tudományos publikációk, mind a vezető szakmai szervezetek és tanácsadók anyagai a mesterséges intelligencia kapcsán figyelembe vesznek további kockázattípusokat, valamint bemutatnak számtalan egyedi kockázatot is. Ennek megfelelően specifikus MI-kockázati területeket jelentenek a magánadat védelméhez kapcsolódó, általános adatbiztonsági, igazságossági, gazdasági és katonai kockázatok.

---

[6] CARDOSO, Pedro Bandeiras – DOMINGOS, Dulce – RESPICIO, Ana: Contributions for risk assessment of IoT-aware business processes at different granularity levels. Procedia Computer Science, Volume 192, 2021. pp. 991–1000.

[7] Control Objectives for Information Technologies. Egy informatikai irányítással foglalkozó nemzetközi szakmai szövetség (Systems Audit and Control Association – ISACA) által létrehozott keretrendszer az IT-menedzsment és -irányítás számára.

[8] CISA Review Manual. https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004KoCbEAK; letöltés: 2022.07.08.

„Hagyományos” gépek és eszközök esetén a biztonság fontos ismérve a balesetek elkerülésének kérdése, amit természetesen az MKI-re is értelmezni kell. Ennek jó példája egy, az önvezető járművekhez kapcsolódó problémák és kihívások rendszerezett vizsgálatát bemutató elemzés,[9] amely a következő területeket érinti: fizikai biztonság, funkcionális/működési megbízhatóság és jogi felelősség. Az első kettő kapcsolódik például az ISO 26262 (funkcionális biztonsággal foglalkozó) szabványhoz,[10] míg Kerrigan a jogi terület egy értékes áttekintését adja amerikai kontextusban,[11] a szakmai oldalt pedig jól képviseli a Google Brain értekezése az MI-balesetekről.[12]

MI-baleset alatt a gyakorlatban alkalmazott MI nem szándékos, de káros viselkedését értjük. Ezek jellemzően gyenge tervezés eredményei. Több munka is kategorizálja az MI-hez köthető balesetek típusait,[13] többnyire annak érdekében, hogy megelőzésükre kutatási irányokat jelöljön ki. A baleset oka lehet hibás célfüggvény, amely mögött azonban állhat szándékos fejlesztői döntés is (például bizonyos ismert negatív mellékhatások elkerülésének szándéka), vagy a jutalom szándékos manipulálása egy bizonyos cél elérése érdekében. De előfordulhat az is, hogy a célfüggvény rendszeres kiértékelése túlzottan erőforrás igényes, és a megtakarítás érdekében nem történik elegendő mélységű betanítás (azaz pl. a felügyelt tanítás nem skálázható[14]).

Szintén előfordulhat nem kívánatos viselkedés már a betanítás során is (ha a fejlesztők például csak a biztonságos helyzeteket kezelik, vagy ha a tanítóhalmaz eloszlása nem felel meg az alkalmazás során várhatónak). Vegyük észre, hogy itt is megfigyelhető a biztonság és a kockázat kettőssége.[15] Széles áttekintést ad az MI-technológiákhoz kapcsolható, a technológia szintjén megjelenő speciális problémákról a Microsoft tanulmánya,[16] amely górcső alá veszi nemcsak a technikai-matematikai hibákat, hanem részletesen taglalja a szándékos támadásokat is. A szándékos támadások között legfontosabbak a bemenet láthatatlan manipulálása a kimenet

[9] TAEIHAGH, Araz – LIM, Hazel Si Min: Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. Transport Reviews, Volume 39, Issue 1, 2019. pp. 103–128.

[10] ISO 26262-1:2011: Road vehicles – Functional safety – Part 1: Vocabulary. https://www.iso.org/standard/43464.html; letöltés: 2022.10.12.

[11] KERRIGAN, Charles: Artificial Intelligence: Law and Regulation. Edward Elgar Publishing, Cheltenham, UK, 2022.

[12] AMODEI, Dario – OLAH, Chris – STEINHARDT, Jacob – CHRISTIANO, Paul – SCHULMAN, John – MANÉ, Dan: Concrete Problems in AI Safety. ArXiv Preprint ArXiv:1606.06565, 2016. https://arxiv.org/pdf/1606.06565.pdf; letöltés: 2022.08.10.

[13] Uo.

[14] STEINHARDT, Jacob: Long-term and short-term challenges to ensuring the safety of AI systems. Academically Interesting blog, 2015.06.24. https://jsteinhardt.wordpress.com/2015/06/24/long-term-and-short-term-challenges-to-ensuring-the-safety-of-ai-systems/; letöltés: 2022.07.12.

[15] MOHSENI, Sina – WANG, Haotao – YU, Zhiding – XIAO, Chaowei – WANG, Zhangyang – YADAWA, Jay: Practical Machine Learning Safety: A Survey and Primer. ArXiv Preprint ArXiv:2106.04823, 2021.

[16] SHANKAR, Ram – KUMAR, Siva – SNOVER, Jeffrey – O'BRIEN, David – ALBERT, Kendra – VILJOEN, Salome: Failure Modes in Machine Learning. Microsoft, November 2019. https://docs.microsoft.com/hu-hu/security/engineering/failure-modes-in-machine-learning; letöltés: 2020.05.21.

félrevezetése érdekében, a modell visszafejtése vagy ellopása, esetleg átprogramozása. Érdemes megemlíteni, hogy a támadásos veszélyek kivédésére a tanítás során a valós ellenséges inputhoz hasonló tanítási bemenet szimulálásával készítik fel az MI-rendszert (pl. neurális hálókat) a veszély kivédésére.[17]

Az MI-szoftverek gyakran válnak nagyobb, meglévő rendszerek részévé, de a beépítendő modellek ellenőrzése és hatásaik felmérése sokkal összetettebb, mint hagyományos alkalmazások esetén, ami növeli a platformokhoz és a beszállítókhoz köthető informatikai kockázatokat. Tehát a mérnöki területekhez hasonlóan az MI-re is érvényes kell legyen a megbízhatóság, a robusztusság, a kockázatérzékenység és a biztonságos tesztelés követelménye,[18] különösen ahogy tovább nő az MI-rendszerek bonyolultsága, ereje, rugalmassága és autonómiája. Ugyanakkor az MI sok szempontból keményebb kihívások elé állítja a fejlesztőket és a felhasználókat, mint más, hagyományos mérnöki alkotások, de akár csak a jelenlegi információrendszerek.

Felmerül a kérdés, miért várhatók magasabb valószínűségű és veszélyesebb kockázatok az MI esetén, mint a hagyományos mérnöki rendszerek, de akár a már meglévő információrendszerek esetén? Az irodalom tipikusan a következő főbb okokat tárgyalja.[19] Mivel az MI képes önálló, autonóm ágensként viselkedni (már egy összetettebb, de még „keskeny" formájában is), ez azt hozza magával, hogy az MI egy jóval nyitottabb problématérben mozog, és a várható viselkedési tere is tágabb, miközben képességei az adott területen meglehetősen erősek is lehetnek. Ennek folyománya, hogy a rendszer viselkedése előre meg nem jósolható lesz, azon belül pedig a helytelen optimalizálás negatív mellékzöngéi felerősödhetnek, kiszámíthatatlanná válhatnak. Mindez úgy, hogy a teljes rendszer maga és viselkedésének részletei át nem láthatók. Továbbá, ha az MI képes tanulni a tapasztalataiból és ha kognitív képességei sokkal gyorsabbak az emberénél, akkor a rendszer képességei nemcsak változhatnak, de e változások lefolyása igen gyors is lehet – ami ellehetetleníti a fejlesztések hagyományos, hibákat idővel kiküszöbölni képes folyamatát. Végül meg kell jegyezni, hogy a jelenlegi gépi tanulási módszerek lehetővé teszik az MI-eszközök számára bizonyos tudáselemek gyors elsajátítását, de nem támogatják a hagyományos emberi értékek beépítését.[20]

Például az MI-modellben esetlegesen rejtőző torzítás (*bias*) beépülhet és rögzülhet az (új) döntési folyamatban.[21] Az adatban rejlő atipikus minták észrevétlenül (és negatívan) befolyásolhatják a döntéseket. Emellett a használat közben feldolgozott adatok mintái eltérőek lehetnek a betanított adatmintáktól, ami a

[17] LI, Guofu – ZHU, Pengjia – LI, Jin – YANG, Zhemin – CAO, Ning – CHEN, Zhiyi: Security Matters: A Survey on Adversarial Machine Learning. ArXiv Preprint ArXiv:1810.07339, 2018. https://www.researchgate.net/publication/328353230_Security_Matters_A_Survey_on_Adversarial_Machine_Learning; letöltés: 2021.11.17.

[18] ORTEGA, Pedro A. – MAINI, Vishal: Building Safe Artificial Intelligence: Specification, Robustness, and Assurance. DeepMind Safety Research, 2018.09.27. https://deepmindsafetyresearch.medium.com/building-safe-artificial-intelligence-52f5f75058f1; letöltés: 2022.12.30.

[19] STEINHARDT, Jacob: Long-term and short-term challenges to ensuring the safety of AI systems. Academically Interesting blog, 2015.06.24.

[20] Uo.

[21] SILBERG, Jake – MANYIKA, James: Tackling bias in artificial intelligence (and in humans). McKinsey and Company, 2019.06.06.

modell teljesítményének romlását eredményezheti (*distribution mismatch*). Ez a jelenség úgy is érzékelhető lehet, hogy a bemeneti adatok mintázata idővel változik (*model aging*). Az adatminőségi problémák különösen magas kockázatú tényezőnek tekinthetők bizonyos nagy érzékenységű területeken, mint például az egészségügy.

A *szervezeti* alkalmazói környezetre vonatkoztatva a McKinsey tanulmánya[22] felsorol modellhez kapcsolható, működési, jogi, hírnevet érintő, továbbá megfelelőségi és egyéb szabályozói (pl. adatvédelmi) kockázatokat. Egyes szervezetekben az MI terjedésével elmosódhatnak a felelősségi határok, vagy akár el is veszhet a számonkérhetőség egyértelműsége, hiszen könnyen elmosódhat, hogy ki a felelős egy adott döntésért, különösen hálózatos önálló ágensek esetén. A másik oldalról pedig felmerülnek az MI-alkalmazás során a szabályozói elvárásoknak való megfelelés kihívásai, a jogi normák esetleges be nem tartásának kockázatai.

Tovább erősítheti a fenti (szervezeti) kockázatokat, hogy az MI-alkalmazások szervezeti határokon is átnyúlhatnak, így a funkciókhoz kötött kockázatmenedzselési megoldások nem hatékonyak. Emellett az MI vezető technológiai óriásai több elemző értékelése szerint[23] igazából államként viselkedő cégek, amelyek olyan határokon átnyúló globális szervezetek (pontosabban szervezeti hálók), amelyek méretes pénztőkével rendelkeznek, de az MI szempontjából még fontosabb, hogy hatalmas adathalmazok, technológiai erőforrások és tudástőke (know-how-k, szabadalmak stb.) felett rendelkeznek.

Ezen túlmenően a közvetlen, az alkalmazó szervezetet, annak üzleti folyamatait és a résztvevőket közvetlenül érintő kockázatokon túl figyelni kell az egyre erősebb MI-rendszerek *hosszú távú,* illetve *társadalmi* következményeire is.[24] A másodlagos vagy akár harmadlagos hatások közül a legtöbbet és legellentmondásosabban a munkaerőpiaci változásokat emelik ki:[25] a munkahelyeket a korábbi technológiai forradalmakhoz képest átfogóbban érintik az MI bevezetésével járó átalakulások. Az MI sokoldalúbban fejti ki hatását ezen a területen, hiszen többféle feladatot és sokféle módon tud átvenni, helyettesíteni, kiegészíteni vagy módosítani. Ebből adódik, hogy az egyes iparágak vagy szakmák nagyon eltérő módon, mélységben és időtávlatban lesznek érintettek[26] az adott terület által igényelt képességek függvényében.

A tágabb társadalmi-politikai következmények közé sorolható többek között a politikai manipuláció lehetősége, a túlzottan felgyorsuló (és egyéni szinten nehezen követhető technológiai innováció és az azzal járó hétköznapi változások, a gazdagság egyenlőtlen eloszlásának további torzulása, vagy akár az MI nem kívánatos (vagy

[22] CHEATHAM, Benjamin – JAVANMARDIAN, Kia – SAMANDARI, Hamid: Confronting the risks of artificial intelligence. McKinsey and Company, 2019.04.26.

[23] ZINGALES, Luigi: Towards a political theory of the firm. Journal of Economic Perspectives, Volume 31, Issue 3, 2017. pp. 113–130.

[24] RUSSELL, Stuart J. – NORVIG, Peter: Artificial intelligence – a modern approach. Prentice Hall, Englewood Cliffs, New Jersey, 2021. https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf; letöltés: 2022.04.17.

[25] VICSEK, Lilla: Artificial intelligence and the future of work–lessons from the sociology of expectations. International Journal of Sociology and Social Policy, Volume 41, Issue 7-8, 2020. pp. 842–861.

[26] PwC: How will AI change the future-of-work? https://www.pwc.com/gx/en/archive/about/analyst-relations/future-of-work.html; letöltés: 2022.12.20.

etikailag megkérdőjelezhető) célokra történő alkalmazása (mint pl. a teljesen autonóm harcászati eszközök vagy a totális megfigyelés). A gazdasági és a piaci kockázatok között érdemes megemlíteni még az adat- vagy technológiai monopóliumok kialakulásának lehetőségét, az esetleges jövedelemtorzulásokat, a fent már említett munkaerőpiaci mozgások közül a munkahelyek nagy arányú megszűnését bizonyos iparágakban vagy piaci szegmensekben. Érdekesség, hogy felmerülhet az emberi intelligencia elértéktelenedése is, hiszen egy esetleges erősebb gépi intelligencia mellett az ember már nem lesz „egyedi" és különleges, a gép mindenben jobb lehet, aláásva egyes csoportok önértékelését.[27]

Egy ritkán felvetett probléma az emberi viszonyok újraértelmezésének kockázata,[28] ami azt jelenti, hogy a technológia átkeretezi az egyén és a szervezet eltérő érdekei között feszülő problémát, és azt úgy állíthatja be, mintha csupán technikai és így könnyen megoldható kérdésről lenne szó. Ezzel egyúttal háttérbe is szorítja az emberi (szociális és gazdasági) szempontokat. Azt is figyelembe kell venni, hogy az emberi szereplőnek nehézséget okozhat az együttműködés a rendszerrel, hogy hogyan integrálja a saját és a rendszer „szakértelmét".[29] Ha nehézségei adódnak, akkor könnyen előfordulhat, hogy megtanulja a rendszer vagy a bemenete manipulálását, hogy számára kedvező eredmények szülessenek.

Az MI-hez kapcsolódó kockázat lehet az is, ha jó megoldások nem terjednek el (pl. az ún. *opt-out* lehetőség megerősödésével[30]), azaz a szervezet vagy társadalom nem részesedik a lehetséges hasznokból és előnyökből – főleg, ha a kockázatok egy elfogadható mérték alatt tarthatók. Ilyen például (a már említett) önvezető járművek helyzete, ahol a veszély kettős: ha túldimenzionálják az előnyöket és a (hétköznapi) gyakorlati tapasztalat mást mutat, például számos baleset történik a bevezetés során, akkor az évekre visszavetheti a fejlődést a tényleges előnyök kárára, míg a kockázatok túlértékelése eleve nehézzé teheti az elfogadást. Ebben a példában a kérdés gyakran úgy merül fel, hogy okozhat-e az önvezető jármű balesetet egyáltalán, azaz az elvárás a teljes, 100%-os biztonság lehet-e. De az igazi kérdés nem az, hogy ez elérhető-e, hanem hogy egy önvezető autó azonos körülmények között kisebb valószínűséggel, vagy enyhébb baleseteket okoz-e, mint ugyanolyan feltételek mentén egy átlagos sofőr. Persze a közvélemény hajlamos eltúlozni egy-egy látványos önvezető fiaskót, különösen, ha az MI tévedése az ember számára elképesztő. Természetesen egy baleset mindig nehéz az érintettek számára, de ettől még lehet, hogy az autonóm jármű – vagy általában az MI – (átlagosan és tipikusan) sokkal jobban teljesít az embernél.

---

[27] RUSSELL, Stuart J. – NORVIG, Peter: Artificial intelligence – a modern approach. Prentice Hall, Englewood Cliffs, New Jersey, 2021.

[28] CURRIE, Morgan – FELDMAN, Jessica – HIMMELREICH, Johannes – NIKER, Fay: Coding Caring Workshop Report. Focused study for the Stanford AI100 Report. Stanford University, 2019. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/coding_caring_workshop_report_1000w_0.pdf; letöltés: 2021.12.21.

[29] MIRBABAIE, Milad – BRÜNKER, Felix – MÖLLMANN, Nicholas R. J. (Frick) – STIEGLITZ, Stefan: The Rise of Artificial Intelligence – Understanding the AI Identity Threat at the Workplace. Electronic Markets, Volume 32, Issue 1, 2022. pp. 73–99.

[30] BONNEFON, Jean-François – SHARIFF, Azim – RAHWAN, Iyad: The moral psychology of AI and the ethical opt-out problem. Oxford University Press, Oxford, UK, 2020. pp. 109–126.

Az MI (és általában a digitális, adatközpontú, hardverigényes IT-megoldások, mint pl. az IoT vagy a kripto) nyersanyag- és energiaéhes technológia, amely így magában hordozza mind a káros környezeti hatások, mind pedig az erőforrások megszerzésének kockázatát.[31]

Az irodalomban tárgyalt MKI-kockázatok áttekintése után fordítsuk figyelmünket egy átfogó keretrendszer alapjai felé.

Ökoszisztéma mint elemzési keret

Az MI-hez kapcsolódó kockázatok a fenti irodalomelemzés szerint tehát több szinten és változatos formában jelenhetnek meg, a technológiai problémáktól az egyéni és társadalmi kihívásokon át a negatív társadalmi-környezeti hatásokig. Nem csoda, hogy a vizsgálatok és a tanulmányok egy-egy részterületre vagy jelenségre fókuszálnak, kevesebb lehetőséget hagyva egy átfogó, integrált elemzésre. A probléma egyik gyökere – a bonyolultságon túl – az egységes vizsgálati keretrendszer hiánya.

Jelen tanulmány javaslatot tesz egy MI-ökoszisztéma kialakítására, amely lehetővé teszi az MI-hez kapcsolódó (és fent áttekintett) kockázatok egységes keretben történő elemzését. Az alkalmazott ökoszisztéma-definíció részben a biológiából ismert (hagyományosnak tekinthető) ökoszisztéma-fogalomra,[32] részben a kilencvenes évek információökológia[33] (IÖ) koncepciójára épít. Az alap ökoszisztéma egy olyan fejlődő, önszerveződő biológiai rendszer, amelyet egy adott fizikai környezetben található organizmusok alkotnak úgy, hogy mind egymással, mind az adott környezettel komplex, dinamikus kapcsolatban állnak összetett visszacsatolási és kiigazító mechanizmusokon keresztül. Az információökológia pedig az emberek, szokások, értékek és a technológia összekapcsolt rendszerét írja le egy lokális környezetben.[34]

Az előbbit gyakran alkalmazzák technológiamenedzsment-kontextusra adaptálva, ha összehasonlítható a fenti definíció szerinti viszonyokkal,[35] míg az utóbbi azért releváns, mert annak egyik alappilére a tudás és a tudásmenedzsment, ami az MI számára is fontos, hiszen az MI-ben a modellezett területre vonatkozó tudás halmozódik fel. Az információökológia mellett külön figyelmet érdemel az adatökoszisztéma[36]

---

[31] OMOHUNDRO, Steve: Autonomous technology and the greater human good. Journal of Experimental & Theoretical Artificial Intelligence, Volume 26, Issue 3, 2014. pp. 303–315.

[32] Ecosystem az OxfordDictionaries.com online szótárban. https://www.oxfordlearnersdictionaries.com/definition/english/ecosystem; letöltés: 2017.10.12.

[33] DAVENPORT, Thomas H. – PRUSAK, Laurence: Information Ecology: Mastering the Information and Knowledge Environment. Oxford University Press, New York, 1997.

[34] NARDI, Bonnie A. – O'DAY, Vicky L.: Information Ecologies: Using Technologies with Heart. MIT Press, Cambridge, 1999.

[35] TSUJIMOTO, Masaharu – KAJIKAWA, Yuya – TOMITA, Junichi – MATSUMOTO, Yoichi: A Review of the Ecosystem Concept – Towards Coherent Ecosystem Design. Technological Forecasting and Social Change, Volume 136, 2018. pp. 49–58.

[36] PARSONS, Mark A.– GODØY, Øystein – LEDREW, Ellsworth – DE BRUIN, Taco F. – DANIS, Bruno – TOMLINSON, Scott – CARLSON, David: A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. Journal of Information Science, Volume 37, Issue 6, 2011. pp. 555–569.

fogalma is, amely a biológiai metaforára építve dolgozza ki az adatokat feldolgozó résztvevők (szereplők) és folyamataik összekapcsolt rendszerét, amely nyílt adatok esetén tipikusan önszerveződő módon épül fel. Érdemes megemlíteni, hogy maga az ökológia tudománya is azokkal a *folyamatokkal* foglalkozik, amelyek befolyásolják egy adott rendszerben az organizmusok eloszlását és mennyiségét, illetve a rendelkezésre álló energia és anyag eloszlását és átalakítását a vizsgált rendszerben.[37]

Ilyen értelemben az MI-ökoszisztéma egy olyan összetett rendszernek tekinthető, ahol a benne részt vevő szereplők összetett folyamatokon keresztül kapcsolódnak egymáshoz, miközben előkészítik és alkalmazzák az MI-t alkotó modelleket és adatokat. Az MI-ökoszisztéma fentiek szerinti részletesebb leírásához e vizsgálat három lépésben jut el: a) leírja az MI mint információrendszer jellemzőit; b) elemzi a kapcsolódó fejlesztési és szervezeti folyamatokat és az abban megjelenő szerepeket; c) felállít egy ökoszisztéma-megközelítésre épülő modellt.

A mesterséges keskeny intelligencia mint információrendszer

Az MKI-t – jellegéből adódóan – szervezeti környezetben elsősorban döntési, döntéstámogatási, illetve ahhoz kapcsolódó, azt megelőző és előkészítő felismerési, értékelési és elemzési folyamatok és feladatok során használják. Jelentős területnek számít a kommunikációs folyamatok kezelése is, ahol a természetes nyelvfeldolgozás egyre összetettebb modelljei állnak a fejlesztések mögött. Az MKI mint döntésben részt vevő ágens esetében a kockázatok megértése és kezelése szempontjából viszont nem az az elsődleges, hogy valamilyen definíció szerint mennyire tekinthető intelligensnek (vagy egyáltalán intelligens-e). Sokkal fontosabb dimenziókat nyit meg a vizsgálat szempontjából, ha az MKI-re információrendszerként tekintünk. Ez a megközelítés lehetővé teszi, hogy megvizsgáljuk a jelenlegi MKI-megoldások alkalmazásának jellegzetességeit, a technológiát, az eszközöket, a fejlesztés és az alkalmazás folyamatát, valamint az abban részt vevő szereplőket – majd erre építve elkészítsük egy MI-ökoszisztéma, azaz egy MI-információökológia modelljét.

De mi is az az információrendszer – amelyet korábban „menedzsmentinformáció-rendszer" (Management Information System – MIS), esetenként „üzletiinformáció-rendszer" (Business Information System – BIS) elnevezéssle is illettek? Alapozó MIS- és BIS-tankönyvekben elég egybehangzóak a különböző definíciók, csak viszonylag apró eltéréseket találni: az információrendszer az emberek, folyamatok, adatok és technológia együttese és kapcsolataik viszonyrendszere. Az emberi és a folyamatoldal szerepelhet szervezetként, míg a technológiát lehet tovább-bontani szoftverekre, hardverekre és hálózati elemekre.[38]

[37] FEDOROWICZ, Jane – GOGAN, Janis L. – RAY, Amy W.: The Ecology of Interorganizational Information Sharing. Journal of International Information Management, Volume 13, Issue 1, 2004. https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1243&context=jitim; letöltés: 2022.09.11.

[38] LAUDON, Kenneth C. – LAUDON, Jane P.: Management information systems: Managing the digital firm. Pearson Education, 2004.

Az MI mint információrendszer elemei e tanulmány javaslata szerint a következők:

- *technológia*: elsősorban szoftver, amely keskeny MI esetén valamilyen (matematikai) modell formájában jelenik meg, és erősen függ az azt felépítő adatoktól. Nagyon speciális és számításigényes célmegoldások (pl. nyelvi vagy játékmodellek) esetén igen erős a hardverigény is, amely lehet költséges dedikált fejlesztés is.

- *adatok*: lehetnek mind történeti, mind valós idejű adatok, numerikus, képi vagy szöveges formában, amelyeket különböző módon tokenizálva használnak fel a modell építésére és melyeknek mennyisége is és a minősége is egyaránt nagy jelentőséggel bír. A modellek specifikussága miatt viszonylag ritka az eltérő típusú adatok keveredése. A minőség nemcsak a hagyományos adatminőségi, használatbeli jellemzőket takarja, hanem azt is, hogy modellépítéskor használt adatok mennyire relevánsak egy új helyzetben, amikor vagy amelyre a modellt alkalmazzák (azaz a modell érzékeny az adatforrásban bekövetkező változásokra).

- *folyamatok*: MI esetén két folyamat-tér találkozásáról van szó, ahol az MI fejlesztési-alkalmazási folyamatai összefonódnak a meglévő szervezeti folyamatokkal – így az MI bevezetése kapcsán az a kérdés is felmerül, hogyan változnak, alakulnak át akár drasztikusan, akár lépésről lépésre az érintett szervezeti folyamatok.

- *az „ember”*: valamennyi, a fejlesztés, az adatkezelés, a bevezetés és az alkalmazás során megjelenő szereplő vagy érintett (MI esetén az alkalmazásba beleértve az eredmények értelmezését is). Ugyanúgy, mint más információrendszer-típusok esetén, az MKI kapcsán is központi kérdés az (új) rendszer mint IT, valamint a bevezetéssel kiváltott változások és hatások menedzselése. Tehát meg kell vizsgálni kik az érintettek, mi az MI mint információrendszer alkalmazásának és a bennfoglaló szervezeti folyamatok átalakulásának a hatása az egyes érintettekre, milyen új szerepek jelennek meg, és mi történik a régiekkel.

Mint említettük, ez az MI-információrendszer két folyamat találkozásánál jelenik meg, így a figyelem most e kettős térre irányul.

Az MI munkafolyamat szakaszai és viszonyuk a szervezeti folyamathoz

Az előző két fejezet logikájából következően az MI-kockázatok szempontjából meg kell tehát vizsgálni az MKI mint információrendszer fejlesztési és alkalmazási folyamatait, e folyamatok szereplőit és kapcsolataikat, továbbá kapcsolódási pontjaikat magával a jellemzően új MI-vel. Kockázatok ugyanis e folyamatok találkozása során léphetnek fel valamely tevékenység, eredmény vagy szereplő kapcsán. Ezt kell összehangolni az alkalmazó szervezeti folyamattal. Erre a részletes vizsgálatra építhető fel egy olyan MI-ökoszisztéma, amely az általános információrendszer-paradigmára támaszkodva, de azon túllépve lehetőséget fog adni a lehetséges kockázatok szisztematikus feltérképezésére, és így a megelőzési és a kezelési lehetőségek keresésére is.

De mit is takar az MI-hez kapcsolódó munkafolyamat? Jelen tanulmányban nemcsak a szigorúan vett kifejlesztés tevékenységeit, hanem az alkalmazás (bevezetés és használat) menedzselési, üzemeltetési és felügyeleti feladatait is figyelembe vesszük. Az általános folyamat természetesen gazdagodhat az adott alkalmazási terület vagy kontextus speciális elemeivel – azaz mindig egy adott tématerületről szól. Az eltérő technikákkal kifejlesztett és különböző kontextusokban alkalmazott MI-rendszerekhez kapcsolódó tényleges folyamatok azonban igen eltérőek lehetnek. Emellett a fejlesztés és a bevezetés is lehet iteratív, illetve számos visszacsatolást is tartalmazhat. Az alkalmazások jelentős része nem teljes mértékben egyedi fejlesztés, hanem valamely meglévő modellre (típusra) épít, azt fejleszti, tanítja, illetve elérhetők részben betanított modellek is – pl. bizonyos jogi szakkifejezésekkel előkészített, online elérhető és testreszabható dokumentumosztályozó MI-rendszerek. Ennek megfelelően két leegyszerűsítést teszünk: egyrészt egy részletes folyamat helyett három fő szakaszt különítünk el, másrészt azon belül tevékenységeket, feladatokat tekintünk – azaz nem szigorú sorrendű vagy elágazásokat tartalmazó konkrét folyamatot. Ez azért is elegendő, mert e dolgozatnak nem célja egy részletes folyamat kidolgozása, a kockázati modell kidolgozása szempontjából a szereplők, illetve viszonyaik és feladataik fontosabbak. A következő átfogó tevékenységlista szolgál megfelelő kiindulópontként az ökoszisztéma felépítésére, amely tevékenységeket négy szakasz mentén mutatunk be: *fejlesztés*, *bevezetés*, *alkalmazás* és *hatások*.

Az MI-munkafolyamat 1. szakasza a *fejlesztés,* amelynek tipikusan három elkülönülő területe van: a modell kidolgozása, az adatok kezelése és az (alap)modell betanítása, tevékenységei pedig a következők:

- modelltípus kiválasztása (részletes tipológiát számos könyv[39] közöl);
- tanítási módszer meghatározása (feltételek, ciklusok stb.);
- bemeneti jellemzők, paraméterek meghatározása, kiválasztása, szűrése;
- kimeneti változó, változók meghatározása;
- szükséges adatok meghatározása (bemenettől is függően);
- adatgyűjtés;
- adattárolás;
- adatok szűrése;
- adat-előkészítés (pontosság biztosítása, torzítás kiszűrése, érzékeny adatok védelme);
- (célzott) modellépítés, tanítás;
- teljesítményértékelési szempontok kiválasztása (a modell jóságának és igazságosságának a meghatározására);
- modell teljesítményének elemzése (benne oksági összefüggések vizsgálata);
- modell finomhangolása;
- utómunkálatok.

[39] RUSSELL, Stuart J. – NORVIG, Peter: Artificial intelligence – a modern approach. Prentice Hall, Englewood Cliffs, New Jersey, 2021.

Mint látható, a tanító adathalmaz megalkotása, konstruálása ugyanolyan fontos, mint magának a modellnek a megépítése, különösen hagyományos rendszerekhez viszonyítva. Az adatgyűjtésnél fontos kérdés, hogy kiről szól, kit érint – egyrészt, hogy milyen demográfiai-szociológiai csoportokról szól, másrészt, hogy milyen viszonyban van a felhasználókkal – azaz általános, külső adatokról van-e szó, vagy az érintett szervezet saját adatairól. Fontos észrevenni, hogy egyrészt egy MI-alkalmazás készülhet általános formában, de jellemző az adott szervezethez egyedi megoldás kifejlesztése is, másrészt a befogadó szervezeti munkafolyamat is több szakaszból és számos tevékenységből, feladatból állhat.

Az MI-munkafolyamat 2. szakasza a *megoldás bevezetése,* amelynek tipikusan három elkülönülő területe van: a modell (szervezeti) folyamatba illesztése, a modell testreszabása és a helyi adatok kezelése az alábbi tevékenységekkel és feladatokkal:

- szervezeti feladat azonosítása;
- a szervezeti folyamat átalakítása;
- az MI telepítése;
- az MI testreszabása, helyi jellemzők betanítása;
- az MI tesztelése;
- az alkalmazottak (dolgozók és IT) betanítása az új, MI-vel gazdagított folyamatra és az MI használatára.

Ahogy az MI egyre jobban elterjed, két vonatkozó jelenség is megfigyelhető: nemcsak egyre több MI-alkalmazást telepítenek különböző szervezeti folyamatokba (jellemzően döntési pontokra), de egyre több beszállítói szoftver is hordoz magában beépített MI-komponenst vagy -megoldást.

Az MI alkalmazási folyamatának 3. szakasza az *MI (-modell) (napi) használata*, amely a következő tevékenységeket foglalja magában:

- alkalmazói adatok gyűjtése, frissítése;
- adatellenőrzés (helyi adatok összehasonlítása a tanításai adatokkal – disztribúciós eltérés kizárása);
- modellalkalmazás (döntésben, döntés-előkészítésben, kommunikációban vagy egyéb feladatra);
- szükség esetén az eredmények értelmezése, szakértői interpretációja;
- modell tanítása alkalmazás közben;
- monitorozás;
- visszajelzés;
- a modell teljesítményének ellenőrzése, nyomon követése;
- modell finomhangolása – adott esetben új jellemzők bevonása (és vissza az első lépésekhez), szélsőséges esetben a kimeneti változó finomítása, akár cseréje.

Érdemes kiemelni annak szükségességét, hogy meg kell érteni a fejlesztő szerepét az adott ökoszisztémában, vagy még pontosabban a fejlesztői szerepeket a specifikus feladatban, amelyben az MI alkalmazását tervezik. Fontos kérdés, hogy részese-e a munkát támogató (vagy az embert kiváltó) MI-technológia (modell) fejlesztője az adott munkafolyamatnak, érti-e a tervezett alkamazás kontextusát,[40] mennyire követi a megoldás alkalmazási tapasztalatait – azaz összességében mennyire válik el az alap, általános modell fejlesztése a később alkalmazásra kerülőtől és a használat során megvalósuló továbbfejlesztéstől, tanulástól. Mindenképp szükséges valamennyi lényeges érintett véleményének a feltérképezése, beleértve azokat, akik alkalmazi fogják, és azokat is, akikre vonatkozik, hatással van (akiket a közvetlen felhasználók közvetve támogatnak).

A fenti szakaszok mellett – azokon túlmutatóan, illetve részben azok eredményeképpen – jelentős irodalom foglalkozik egy fontos kérdéssel, az MI bevezetésének és használatának *hatásaival* – mind a szervezeten belül, mind azon kívül. Természetesen ez nem a szervezeti munka értelmében vett szakasz, hanem az MI-rendszer által hozott változások és eredmények utórezgéseit, a szervezetre, a szereplőkre, vagy akár a társadalomra gyakorolt hatását igyekszik figyelemmel kísérni, megérteni. Azért is érdemes ezt figyelembe venni, mert a lehetséges hatások – és kockázatok – előzetes felmérése az MI esetén még fontosabb, mint hagyományos szoftvereknél. Legfontosabb tevékenység az érintettek feltérképezése, az alkalmazási folyamatban elfoglalt szerepükkel és az MI-rendszerhez fűződő viszonyukkal együtt: kik és milyen módon érintettek vagy tudják befolyásolni a rendszert és hatásait, a felhasználók munkája hogyan változik; a megbízók elérték-e céljaikat; mit tapasztalnak azok, akikre az MI által támogatott (vagy hozott) döntések vonatkoznak; van-e ellenséges beavatkozás, manipuláció.

A mesterséges intelligencia szervezeti folyamatainak szereplői

A fenti szakaszok és tevékenységek segítségével feltérképezhetők azok a szerepek, amelyek részt vesznek a feladatok végrehajtásában és az MI szervezeti folyamatba építésében vagy hatással vannak ezekre:

- *adatszolgáltatók*: jellemzően passzív, esetleg aktív szereplők és érintettek, akikről az adatokat gyűjtik;
- *adatkezelők*: akik az adatokat gyűjtik (sokszor automatikusan, különböző módon és forrásból), kezelik, tárolják, rendezik, tisztítják, esetleg kategorizálják, előkészítik stb.
- *MI-eszközfejlesztők*: azoknak a (alap)modelleknek, eszközöknek, modellező környezeteknek a kidolgozásával foglalkoznak, amelyek mintegy építőkockaként használhatók konkrét (eseti) modellek kifejlesztésében, beleértve a kapcsolódó platformelemeket is;

[40] CURRIE, Morgan – FELDMAN, Jessica – HIMMELREICH, Johannes – NIKER, Fay: Coding Caring Workshop Report. Focused study for the Stanford AI100 Report.

- *adattudósok* (*Data Scientists*): a modell építésével és betanításával foglalkoznak – a modelltípus vagy -típusok kiválasztásával, felépítésével, paraméterek és proxyk definiálásával, a modell tanításával és finomhangolásával;
- *üzemeltetés*: háttér IT biztosítása;
- *törvényalkotók és szakmai szervezetek*: akik a környezetet (akár a szabályozói, akár a szakmai közvéleményt) meghatározzák vagy az ipari kontextusra hatással vannak;
- *felhasználók*: akik az MI-eszközt munkájuk során ténylegesen használják – jellemzően döntéshozásban vagy valamely szervezeti folyamat, lépés automatizálásában;
    - az érintettségük jellegénél fogva külön érdemes kezelni azokat az alkalmazottakat, akiket az MI esetlegesen helyettesít, hisz „szerepük" értelmében ők nem felhasználók, hanem más módon lesznek érintettek;
- *értékeltek*: akiknek az adatait az MI alkalmazása során használják – és akikről a kimenet szól;
    - szervezeti döntések esetén az alkalmazás adatai vonatkozhatnak valakire vagy valakikre, akikről a döntés szól: a vizsgálat vagy kiértékelés tárgya lehet konkrét személy vagy intézmény;
- *szakértők*: lehetnek szakmai (az adott alkalmazói szakterületről) vagy általános MI-szakértők, akik jellemzően az MI, a modell eredményeinek az értelmezését segítik, amennyiben arra szükség van;
- *megbízók*: akik a modell alkalmazását elrendelték és jellemzően fizetnek érte, vagy valamilyen módon felelősek a bevezetésért (de nem közvetlen felhasználók);
- „*rosszakarók*": az MI rosszindulatú befolyásolásában, kihasználásában, rongálásában, manipulálásában, kihasználásában stb. érdekelt szereplők, akik lehetnek külső vagy belső viszonylatban az alkalmazó szervezethez képest;
    - ez a szerep természetesen egy általános kategóriát jelöl és sokféleképp specifikálható egy adott konkrét kontextusban, illetve bármely más szereppel összekapcsolható (azaz bármely adott szerepet felvevő konkrét szereplő ezt a szerepet is felveheti);
- *hatásban érintettek*: akikre az MI és a modell alkalmazása, illetve a kimenetből következő eredmények, döntések hatással vannak;
    - alapból egybeesik az *értékeltekkel*, de a kör tágabb, hiszen lehetnek a szervezet saját dolgozói (pl. akiknek a munkáját érinti), külső szereplők, vagy általában a társadalom, de az érintettek kiléte nem mindig állapítható meg közvetlenül.

Néhány fontos megjegyzés a fenti szerepekhez:

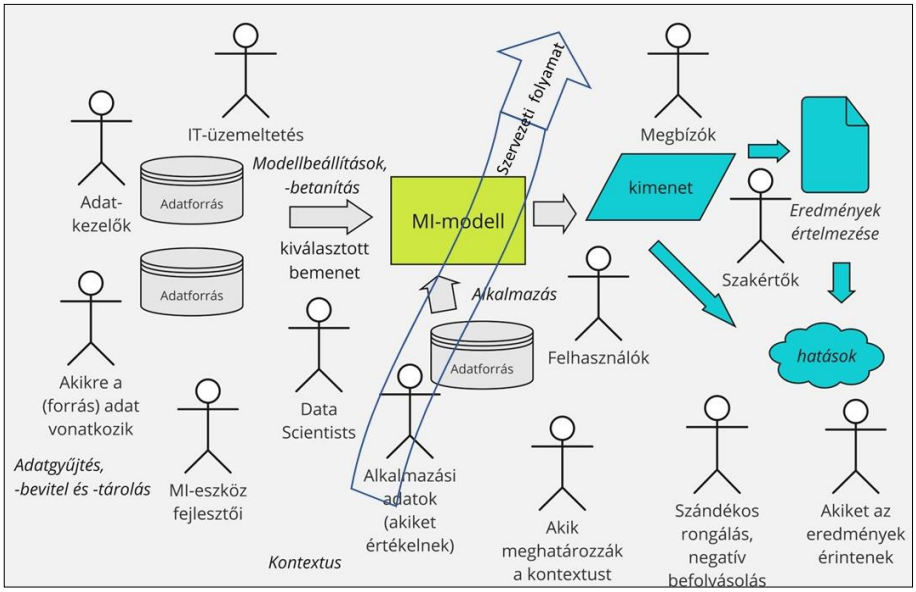
- ezek a (sztereotipikusnak tekinthető) szerepek egy adott helyzetben, konkrét fejlesztési és szervezeti folyamat mentén válnak egyértelművé;
- lehetnek igen összetettek is, illetve további, finomabb szerepekre is bonthatók;
- nem minden szerep fordul elő minden helyzetben;
- egy adott szerepet konkrét esetben felvehet egyén, csoport, szervezet (pl. kormány) is;
- egy (konkrét) szereplő több szerepkörben is megjelenhet.

Az így megismert szerepek elhelyezhetők a már említett kettős folyamattérben.

Mesterségesintelligencia-ökoszisztéma modell

Mint az ökoszisztéma definícióiból kiderül, a fókusz a szereplőkön és kapcsolataikon, azaz a rendszerben elfoglalt helyükön van. A fent taglalt folyamatok és szerepek viszonyában ennek megfelelően egy összetett ökoszisztéma alakul ki, ahol a kapcsolatok a folyamat(ok) mentén értelmeződnek és a meglévő (jellemzően döntési vagy döntési pontokat tartalmazó) szervezeti folyamatba integrálódik az MI-fejlesztési-alkalmazói folyamat, így megváltoztatja az egyes döntések lefolyását vagy akár a folyamat egészét. Ez a változás lehet az emberi döntéshozó támogatása (pl. döntés-előkészítés formájában), az emberi munka kiegészítése (augmentációja), de akár a folyamat teljes automatizációja is (melynek egyik formája a Robotic Process Automation).

Az 1. ábrán látható vázlatos ökoszisztéma-modell általános szerepeket és folyamatelemeket ír le. Valós alkalmazása egy konkrét helyzetben két lépésben történik. Előbb a szerepek és a folyamatok részletes kibontása szükséges egy adott iparágra vagy szakterületre. Ekkor az egyes szerepek a vizsgált területtől függően további részszerepekre bonthatók (ahogy fent is említettük). Az így finomított modellt lehet konkretizálni egy specifikus szervezet vagy célzott szakmai alkalmazó közösség szintjén. Ezen a harmadik szinten a szerepeket konkrét, nevesített szereplők is felvehetik, vagy a szerep specifikálható az adott szűkebb környezetre – és természetesen a folyamatok és az adatok is konkrétan behelyettesíthetők. Az általános modell nem jelzi a lehetséges visszacsatolásokat, hiszen azok nem állapíthatók meg azon a szinten, felismerésük és elemzésük szintén a harmadik szinten (a második lépés során) válik lehetségessé. Ebből a megközelítésből adódik, hogy vagy iparági vagy szervezeti mélységben lehet csak felismerni, ha valamely szerepeket ugyanaz a szereplő (vagy csoport) tölti be – amely esetben felmerülhet az önbezáró és önerősítő visszacsatolás veszélye vagy a manipuláció lehetősége.



1. ábra. A mesterséges keskeny intelligencia ökoszisztéma-modelljének vázlata
Szerkesztette: Csáki Csaba

MI-kockázatkategóriák és MI kockázati mátrix

Az ökoszisztéma-modell elemzésénél – figyelembe véve a kockázatról általában és specifikusan az MI-re vonatkozóan korábban (pl. a kockázatok definiálása során) leírtakat – három dimenzió mentén lehet feltérképezni a kapcsolódó kockázatokat: a kockázat oka, (elsődleges) megjelenési helye és a (másodlagos) hatása szerint. A kockázat oka (vagy más értelmezésben megjelenésének típusa) lehet fejlesztői vagy felhasználói *hiba*, sérülékenység (mely az MI mint rendszer *sajátosságaiból* adódik) vagy szándékos *támadás*. A hibák és a támadások szerepekhez (majd a 3. szinten konkrét szereplőkhöz) köthetők, míg sérülékenység alatt az MI természetéből adódó problémákat értünk A kockázat megjelenési helye lehet az adat vagy a modell (azaz elsődlegesen a technológia), míg a kimenetek másodlagos hatásai (melyek túlmutatnak a technológián és az érintett folyamatokon) jelentkezhetek az egyén és a szervezet, vagy az egyén és a társadalom viszonylatában. Az irodalom alapján korábban elemzett kockázatok így egy mátrixba rendezhetők, bát a másodlagos hatásokat érdemes külön is vizsgálni.

A mátrix értelmezéséhez és használatához szintén a korábban (a modellhez kapcsolódóan leírt) háromszintű elemzés alkalmazható annak érdekében, hogy kideríthető legyen, mely kockázatok léphetnek fel nagyobb valószínűséggel egy adott specifikus iparági szituációban vagy konkrét szervezeti MI-alkalmazás esetén.

KMI-kockázatok			Hibák	Az MI természetéből adódó gyengeség	Támadások
Megjelenési helye	Technológia	Adat	Torzított adat (bias) Nem tiszta adatok Nem reprezentatív tanítóadatok Limitált vagy túl nagy adathalmaz Hibás adatátvitel korábbi modellből Adatfrissítés hiánya	Félrevezető énkép az adatokban Alternatív lehetőségek hiánya A méret, a sebesség vagy a komplexitás nem megfelelő kezelése	Az adat szemetelése Tanítóadatok visszafejtése Konkrét személyazonosság visszafejtése (*membership inference*)
		Modell	Nem megfelelő hasznossági függvény Nem a megfelelő szakértői tudás leképezése Hibás modellátvitel Modellfrissítés hiánya Nem megfelelő proxyk alkalmazása	Magyarázhatóság hiánya Átláthatóság hiánya Az érintettek (etikai) preferenciáinak félreértése	A modell visszafejtése (*reverse engineering*) Támadásos módosítás Célzott félreosztályozás Modell másolása, ellopása Modell átverése
	Szervezet		Hibásan célzott vagy rosszul feltett kérdés Nem megfelelő kontextusban történő alkalmazás Hibás feltételezések (mit jelent a szevezetnek)	Arrogáns algoritmus Átlag használó nem érti	Szándékosan hibás MI Valós cél elfedése
Hatás (másodlagos)		Egyén és szervezet viszonyában	Ellenőrzés elvesztése a rendszer felett Szabályozásnak nem megfelelő alkalmazás	Nem tervezett mellékhatás Autonómiával járó (felelősségi) kockázat Felelősségre vonhatóság elvesztése Befolyásolt viselkedés	A megbízhatóság aláásása A rendszerbe vetett bizalom megingatása
		Egyén és társadalom viszonyában	Túl gyors technológiai innováció Az emberek elveszítik speciális helyzetüket és nem érzik magukat hasznosnak (MI jobb) Hibás feltételezések (mit gondolnak róla) Az MI alkalmazása veszélyes vagy káros célra Elszabadult autonóm fegyverek Téves szabályozás	Szociális-társadalmi elszigetelődés Morális relativizmus Istent játszani Munkahelyek elvesztése Egyenlőtlen jövedelmek Növekvő gazdasági különbségek Társadalmi feszültségek a munkahelyek átalakulása miatt	Manipuláció A megbízhatóság aláásása Teljes társadalmi ellenőrzés Piaci monopóliumok kialakulása

1. táblázat. MKI-kockázatok dimenzionált áttekintése
Szerkesztette: Csáki Csaba

MI-kockázatok kezelése: lehetőségek és buktatók

A rendszer fejlesztőinek nemcsak a modellt és az adatokat kell tudniuk „uralni" önmagukban, de fel kell mérniük a használat közbeni várható viselkedést is. Mi több, ezt minden egyes speciális alkalmazási kontextusban meg kell érteniük, külön kitérve arra, várhatóan hogyan fogják használni a rendszert, illetve az adott kontextusban dolgozó jövőbeli felhasználók hogyan fogják interpretálni a rendszer előrejelzéseit vagy javaslatait („döntését"), és hogyan reagálnak azokra. Ez természetesen nem kiszámítható, nincs egzakt tudományos módszer rá – tapasztalat és körültekintés szükséges. A kezelési lehetőségek áttekintését öt irányból érdemes megközelíteni: technikai, szakmai, szervezeti, jogi és társadalmi.

- *Technikai, technológiai megoldások:*
  - beépített szabályozás (*regulation by design*);
  - kontrollált platformok biztosítása;
  - a modell teljesítményének folyamatos vagy rendszeres ellenőrzése.
- *Szervezeti feladatok:*
  - átláthatóság biztosítása;
  - etikai kódex készítése;
  - etikai tanács felállítása.
- *Szakmai utak:*
  - szakértői testületek, szakmai közösségek/szervezetek létrehozása;
  - legjobb gyakorlatok, iránymutatások (lásd pl. IEEE);
  - xMI- és megbízható MI-előírások.
- *Jogi lehetőségek*:
  - szabályozás(ok);
  - szabályozói szervezetek és infrastruktúra.
- *Társadalmi lépések:*
  - informálás;
  - oktatás, felkészítés.

A fenti lehetőségek természetesen kombinálhatók és együttes hatásuk számos kockázat megelőzését és kezelését teheti hatékonyabbá és sikeresebbé.

Összegzés és további kutatási irányok

A fentiekből következik, hogy nagyon nehéz előre megmondani, milyen hatásokkal jár az ember–MI szimbiózisa, és milyen nem kívánatos reakciók és viselkedésformák alakulhatnak ki. Nem tudjuk pontosan előrejelezni a pszichológiai reakciókat és a szociológiai dinamikát. Továbbá számolni kell azzal, hogy egy adott rendszer bevezetése során és következményeként számos módon lehet torzítani a bevezetést (annak célját, hasznát és indokoltságát) vagy a használat módját, de előfordulhat ellenállás és manipuláció is. Arra is fel kell hívni a figyelmet, hogy a mai globális, többdimenziós piacterek, hatalmas adattömegek és növekvő komplexitás olyan típusú kihívások elé állítja e piacok szereplőit, amelyekre a hagyományos (e tanulmányban is említett) kockázatkezelési technikák és a megszokott szervezetirányítási gyakorlatok nincsenek felkészülve és nem adnak megfelelő válaszokat. Ez önmagában is kockázat, ráadásul olyan, amely sokszor rejtve marad, tipikusan nincs a gondolkodás előterében. Az MI fejlesztése és különösen alkalmazása során számos logikai visszacsatolást érhetünk tetten, ami növeli a veszélyes megerősítés lehetőségét.[41]

Végül – amit az irodalom az integrált megközelítés hiánya miatt jellemzően elhanyagol vagy elsiklik fölötte – az a megfigyelés, hogy a jelenlegi MI-rendszerek egyáltalán nem olyan intelligensek, mint amilyennek sokszor gondoljuk őket. Sőt, bár egyes technikai, mintafelismerésen alapuló tevékenységekben messze túlteljesítik az emberi képességeket, a legegyszerűbb helyzetekben is képesek döbbenetesen „buta" dolgokat elkövetni. Ez egyrészt azért meggondolandó, mivel ezek a rendszerek viharos gyorsasággal terjednek és már ott vannak szinte mindenütt, másrészt valós képességeikhez képest túl nagy felelősségű tevékenységekben is lehet irányító, meghatározó szerepük. Talán mégis azoknak lesz igaza, akik komoly veszélyekre figyelmeztetnek – de nem a szuperintelligencia okán, hanem a láthatatlan buta intelligenciák tömege miatt.

E kimenet megelőzésének érdekében a kutatás legfőbb eredménye egy ökoszisztéma-alapú MI kockázati keretrendszer kialakítása az alábbiak szerint.

A javaslat a keskeny MI fejlesztési folyamatának négy szakaszát különbözteti meg:

- fejlesztés: a modell kidolgozása, adatok kezelése és az (alap)modell betanítása;
- a megoldás bevezetése: a modell (szervezeti) folyamatba illesztése, a modell testreszabása és a helyi adatok kezelése;
- az MI (modell) (napi) használata;
- az MI bevezetésének és használatának hatásai (mind az adott szervezeten belül, mind azon kívül).

[41] O'NEIL, Cathy: Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, New York, 2016.

Az MI-munkafolyamatban több új szerep ismerhető fel a régiek mellett. Ezek közül a legfontosabbak: adatszolgáltatók, adatkezelők, MI-eszközfejlesztők, adattudósok, IT-üzemeltetés, kontextust meghatározók (törvényalkotók és szakmai szervezetek), felhasználók (és kiváltott alkalmazottak), értékeltek, szakértők, megbízók, „rosszakarók" és a hatásokban érintettek.

A tanulmány a feldolgozott irodalom és szakmai anyagok (weboldalak és ipari tanulmányok) által felsorolt kockázatok közül ötvenhetet helyez el egy MI kockázati mátrixban.

Mind az ökoszisztéma-modell, mind a kockázati mátrix két lépésben alkalmazható a gyakorlatban: 1) előbb a szerepek és a folyamatok részletes kibontása szükséges egy adott iparágra vagy szakterületre – ekkor az egyes szerepek további részszerepekre bonthatók (ez a 2. szint); 2) az így finomított modellt lehet konkretizálni egy specifikus szervezet vagy célzott szakmai alkalmazó közösség szintjén – ahol a szerepeket konkrét, nevesített szereplők is felvehetik, vagy a szerep specifikálható az adott szűkebb környezetre, illetve a folyamatok és az adatok is konkrétan behelyettesíthetők (ez a 3. szint). Az általános modell nem jelzi a lehetséges visszacsatolásokat, hiszen azok nem állapíthatók meg azon a szinten, felismerésük és elemzésük szintén a harmadik szinten (azaz a második lépés után) válik lehetségessé.

Jelen dolgozat és javasalt szervezeti MI-re koncentrál (tehát pl. nem vizsgálja játékok vagy egyénileg, saját szórakoztatásra fejlesztett MI-programok kockázatait).

A későbbiekben megvalósítható a teljes kutatási program tématerületi csoportjaival együttműködve az egyes területek (ipar, pénzügy, oktatás, eü. stb.) eredményeinek összevetése és elemzése az ökoszisztéma-modell mentén (részben annak validálására), illetve az egyes tématerületekhez kapcsolható megelőzési lehetőségek és feltételeik feltérképezése (pl. szabályozási, fejlesztői-technikai, alkalmazási dimenziók mentén). További lehetséges kutatási feladat az ökoszisztéma-modell és a kockázati mátrix gyakorlati alkalmazása. Az egyes feldolgozott iparágak kockázati térképén (lásd 1. lépés, 2. szint) túl részletes esettanulmányokat is érdemes kidolgozni az irodalomban feldolgozott esetek segítségével (lásd 2. lépés, 3. szint).

IRODALOMJEGYZÉK

AMODEI, Dario – OLAH, Chris – STEINHARDT, Jacob – CHRISTIANO, Paul – SCHULMAN, John – MANÉ, Dan: Concrete Problems in AI Safety. ArXiv Preprint ArXiv:1606.06565, 2016. https://arxiv.org/pdf/1606.06565.pdf; letöltés: 2022.08.10.

BONNEFON, Jean-François – SHARIFF, Azim – RAHWAN, Iyad: The moral psychology of AI and the ethical opt-out problem. Oxford University Press, Oxford, UK, 2020. pp. 109–126.

CARDOSO, Pedro Bandeiras – DOMINGOS, Dulce – RESPICIO, Ana:
Contributions for risk assessment of IoT-aware business processes at different granularity levels. Procedia Computer Science, Volume 192, 2021. pp. 991–1000.
https://www.researchgate.net/publication/355025896_Contributions_for_risk_assessment_of_IoT-aware_business_processes_at_different_granularity_levels; letöltés: 2022.04.27.

CHEATHAM, Benjamin – JAVANMARDIAN, Kia – SAMANDARI, Hamid: Confronting the risks of artificial intelligence. McKinsey and Company, 2019.04.26. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence; letöltés: 2021.11.30.

CISA Review Manual
https://store.isaca.org/s/store#/store/browse/detail/a2S4w000004KoCbEAK;
letöltés: 2022.07.08.

CURRIE, Morgan – FELDMAN, Jessica – HIMMELREICH, Johannes – NIKER, Fay: Coding Caring Workshop Report. Focused study for the Stanford AI100 Report. Stanford University, 2019. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/coding_caring_workshop_report_1000w_0.pdf; letöltés: 2021.12.21.

DAVENPORT, Thomas H. – PRUSAK, Laurence: Information Ecology: Mastering the Information and Knowledge Environment. Oxford University Press, New York, 1997.

Ecosystem az OxfordDictionaries.com online szótárban.
https://www.oxfordlearnersdictionaries.com/definition/english/ecosystem;
letöltés: 2017.10.12.

FEDOROWICZ, Jane – GOGAN, Janis L. – RAY, Amy W.: The Ecology of Interorganizational Information Sharing. Journal of International Information Management, Volume 13, Issue 1, 2004.
https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1243&context=jitim;
letöltés: 2022.09.11.

GASPER, Des – GÓMEZ, Oscar A.: Human security thinking in practice:‘personal security’, ‘citizen security' and comprehensive mappings.
Contemporary Politics, Volume 21, Issue 1, 2015. pp. 100–116.
https://www.researchgate.net/publication/271224384_Human_security_thinking_in_practice_%27personal_security%27_%27citizen_security%27_and_comprehensive_mappings;
letöltés: 2022.04.25.

ISMAGILOVA, Elvira – HUGHES, Laurie – RANA, Nripendra P. – DWIVEDI, Yogesh K.: Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. Information Systems Frontiers, Volume 24, Issue 2, 2022. pp. 393–414.
https://www.researchgate.net/publication/343123843_Security_Privacy_and_Risks_Within_Smart_Cities_Literature_Review_and_Development_of_a_Smart_City_Interaction_Framework; letötlés: 2022.04.25.

ISO 26262-1:2011: Road vehicles – Functional safety – Part 1: Vocabulary.
https://www.iso.org/standard/43464.html; letöltés: 2022.10.12.

KERRIGAN, Charles: Artificial Intelligence: Law and Regulation.
Edward Elgar Publishing, Cheltenham, UK, 2022.

LAUDON, Kenneth C.– LAUDON, Jane P.: Management information systems: Managing the digital firm. Pearson Education, 2004.
https://repository.dinus.ac.id/docs/ajar/Kenneth_C.Laudon,Jane_P_.Laudon_-_Management_Information_Sysrem_13th_Edition_.pdf; letöltés: 2022.10.16.

LI, Guofu – ZHU, Pengjia – LI, Jin – YANG, Zhemin – CAO, Ning – CHEN, Zhiyi: Security Matters: A Survey on Adversarial Machine Learning. ArXiv Preprint ArXiv:1810.07339, 2018. https://www.researchgate.net/publication/328353230_Security_Matters_A_Survey_on_Adversarial_Machine_Learning; letöltés: 2021.11.17.

MIRBABAIE, Milad – BRÜNKER, Felix – MÖLLMANN, Nicholas R. J. (Frick) – STIEGLITZ, Stefan: The Rise of Artificial Intelligence – Understanding the AI Identity Threat at the Workplace. Electronic Markets, Volume 32, Issue 1, 2022. pp. 73–99. https://www.researchgate.net/publication/355094976_The_rise_of_artificial_intelligence_-_understanding_the_AI_identity_threat_at_the_workplace; letöltés: 2022.07.08.

MOHSENI, Sina – WANG, Haotao – YU, Zhiding – XIAO, Chaowei – WANG, Zhangyang – YADAWA, Jay: Practical Machine Learning Safety: A Survey and Primer. ArXiv Preprint ArXiv:2106.04823, 2021.

NARDI, Bonnie A. – O'DAY, Vicky L.: Information Ecologies: Using Technologies with Heart. MIT Press, Cambridge, 1999.

O'NEIL, Cathy: Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, New York, 2016. https://edisciplinas.usp.br/pluginfile.php/4605464/mod_resource/content/1/%28FFLCH%29%20LIVRO%20Weapons%20of%20Math%20Destruction%20-%20Cathy%20ONeal.pdf; letöltés: 2021.12.21.

ØIEN, Knut: A framework for the establishment of organizational risk indicators. Reliability Engineering & System Safety, Volume 74, Issue 2, 2001. pp. 147–167.

OMOHUNDRO, Steve: Autonomous technology and the greater human good. Journal of Experimental & Theoretical Artificial Intelligence, Volume 26, Issue 3, 2014. pp. 303–315. https://pdfs.semanticscholar.org/2f5f/3234fa085ba16fdbbc92527f0759fc9e8d16.pdf; letöltés: 2022.05.24.

ORTEGA, Pedro A. – MAINI, Vishal: Building Safe Artificial Intelligence: Specification, Robustness, and Assurance. DeepMind Safety Research, 2018.09.27. https://deepmindsafetyresearch.medium.com/building-safe-artificial-intelligence-52f5f75058f1; letöltés: 2022.12.30.

PARSONS, Mark A.– GODØY, Øystein – LEDREW, Ellsworth – DE BRUIN, Taco F. – DANIS, Bruno – TOMLINSON, Scott – CARLSON, David: A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. Journal of Information Science, Volume 37, Issue 6, 2011. pp. 555–569. https://journals.sagepub.com/doi/pdf/10.1177/0165551511412705; letöltés: 2021.11.15.

PENCE, Justin – MOHAGHEGH, Zahra – OSTROFF, Cheri – LEE, Ernie – YILMAZ, Fatima – GRANTOM, Rick – JOHNSON, David: Toward monitoring organizational safety indicators by integrating probabilistic risk assessment, socio-technical systems theory, and big data analytics. 12th Int. Probabilistic Safety Assessment and Management Conference, 2014. pp. 237–251. https://www.iapsam.org/psam12/proceedings/paper/paper_549_1.pdf; letöltés: 2022.02.11.

PwC: How will AI change the future-of-work?
https://www.pwc.com/gx/en/archive/about/analyst-relations/future-of-work.html;
letöltés: 2022.12.20.

RUSSELL, Stuart J. – NORVIG, Peter: Artificial intelligence – a modern approach.
Prentice Hall, Englewood Cliffs, New Jersey, 2021.
https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf;
letöltés: 2022.04.17.

SHANKAR, Ram – KUMAR, Siva – SNOVER, Jeffrey – O'BRIEN, David – ALBERT, Kendra – VILJOEN, Salome: Failure Modes in Machine Learning. Microsoft, November 2019.
https://docs.microsoft.com/hu-hu/security/engineering/failure-modes-in-machine-learning;
letöltés: 2020.05.21.

SILBERG, Jake – MANYIKA, James: Tackling bias in artificial intelligence (and in humans).
McKinsey and Company, 2019.06.06.
https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans; letöltés: 2021.12.22.

STEINHARDT, Jacob: Long-term and short-term challenges to ensuring the safety of AI systems. Academically Interesting blog, 2015.06.24.
https://jsteinhardt.wordpress.com/2015/06/24/long-term-and-short-term-challenges-to-ensuring-the-safety-of-ai-systems/; letöltés: 2022.07.12.

TAEIHAGH, Araz – LIM, Hazel Si Min: Governing autonomous vehicles:
Emerging responses for safety, liability, privacy, cybersecurity, and industry risks.
Transport Reviews, Volume 39, Issue 1, 2019. pp. 103–128.
https://www.tandfonline.com/doi/epdf/10.1080/01441647.2018.1494640?src=getftr;
letöltés: 2022.09.04.

TSUJIMOTO, Masaharu – KAJIKAWA, Yuya – TOMITA, Junichi – MATSUMOTO, Yoichi:
A Review of the Ecosystem Concept – Towards Coherent Ecosystem Design.
Technological Forecasting and Social Change, Volume 136, 2018. pp. 49–58.
https://www.researchgate.net/publication/318449271_A_review_of_the_ecosystem_concept_-_Towards_coherent_ecosystem_design; letöltés: 2022.07.08.

VICSEK, Lilla: Artificial intelligence and the future of work–lessons from the sociology of expectations. International Journal of Sociology and Social Policy, Volume 41, Issue 7-8, 2020. pp. 842–861.
https://unipub.lib.uni-corvinus.hu/7075/1/Lilla_Vicsek_2021IJSSP.pdf;
letöltés: 2022.10.21.

ZINGALES, Luigi: Towards a political theory of the firm. Journal of Economic Perspectives, Volume 31, Issue 3, 2017. pp. 113–130.
https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.31.3.113; letöltés: 2022.09.17.