GAZDASÁG ÉS PÉNZÜGY • 12. ÉVF. 1. SZ. • 2025. MÁRC. • DOI: 10.33926/GP.2025.1.5

A MUNKAHELYI MI–HASZNÁLAT OKOZTA MŰKÖDÉSI KOCKÁZATOK ÉS KEZELÉSÜK A SZAKIRODALOMBAN

Juhász Péter[1]

ABSZTRAKT

A mesterséges intelligencia (MI) egyre szélesebb körű munkahelyi alkalmazása jelentős működési kockázatokat hordoz magában, amelyek technikai, etikai, jogi és szervezeti szinten egyaránt megjelennek. A kockázatok többek között az adatvédelmi aggályok, döntéshozatali torzulások formájában mutatkoznak meg, de lefedik a felelősség elmosódását és az emberi munkahelyek átalakulását is. A cikk széles körű szakirodalmi áttekintés alapján rendszerezi az MI-specifikus működési kockázatokat, különös tekintettel az etikai dilemmákra, a szabályozási kihívásokra és a szervezeti hatékonyságra gyakorolt hatásokra. Az elemzés rávilágít az MI által generált paradoxonokra, amelyek bonyolítják a vállalati döntéshozatalt és a kockázatkezelést. A tanulmány végül javaslatokat fogalmaz meg az MI-hez kapcsolódó kockázatok proaktív kezelésére: ezek közül a legfontosabbak: az átláthatóság fokozása, az adaptív kockázatkezelési modellek alkalmazása és a szabályozói környezet folyamatos fejlesztése. Az eredmények azt mutatják, hogy az MI vállalati integrációja nem csupán technológiai, hanem stratégiai és szervezeti kérdés is, amely átfogó, hosszú távú megközelítést igényel.

JEL-kódok: G28, G32, M15, M51, O33, O34

Kulcsszavak: mesterséges intelligencia, kockázatkezelés, szervezeti működés, alkalmazkodás, emberi erőforrás

1. BEVEZETÉS

Bár a mesterséges intelligencia (MI, vagy artificial intelligence – AI) története már több évtizedes, az AI-alapú alkalmazások robbanásszerű terjedése mégisa 2010-es évek elején, a mélytanulás korábbi gépi tanulási technikáját felülmúló hatékonyságával kezdődött (Krizhevsky–Sutskever–Hinton, 2012). Ez az új technológia komoly lendületet adott a természetes nyelvfeldolgozásnak (NLP), miköz-

1 *Juhász Péter* PhD, CFA, habilitált docens, Budapesti Corvinus Egyetem, Pénzügy Intézet. E-mail: peter.juhasz@uni-corvinus.hu.

ben a számítógépek számítási teljesítménye radikálisan növekedett. Ez különösen a grafikus feldolgozóegységek (GPU-k) fejlesztésében jelent meg látványosan, miután az egyesült államokbeli NVIDIA erre a területre kezdett koncentrálni. Az MI-specifikus chipek, például a Google Tensor Processing Unitjai (TPU-k), tovább gyorsították a fejlődést, hiszen nagyméretű neurális hálózatok hatékony létrehozását tették lehetővé.

A fejlődés azonban aligha mehetett volna végbe, ha nem áll rendelkezésre a tanításhoz hatalmas mennyiségű adat, amelyet az internet, a közösségi média és az üzleti alkalmazások generáltak a korábbi évtizedekben. Mindeközben a vállalatok hatalmas számítási kapacitásokhoz fértek hozzá a felhőszolgáltatók (AWS, Google Cloud, Microsoft Azure) révén anélkül, hogy saját infrastruktúrát építettek volna ki.

A lavinát alighanem a Google indította el, amikor 2018-ban bemutatta a BERT (Bidirectional Encoder Representations from Transformers) modellt, amely jelentősen javította a fordítási és szöveggenerálási feladatok pontosságát (Devlin et al., 2018). De az igazi áttörést talán mégis az OpenAI GPT-3 modelljének 2020-as debütálása jelentette. Az egyre fejlődő szöveggenerálási képesség a modellek széles körű alkalmazáshoz vezetett: megjelentek az MI-alapú chatbotok, a virtuális asszisztensek és a tartalomkészítők.

A piacra sorra léptek be az olyan óriások, mint a Google (DeepMind: Gemini), a Facebookot működtető Meta (Llama), az Anthropic (Claude) és a részben az Nvidia és Jeff Bezos finanszírozásával létrejött, de elvileg független Perplexity AI. A legutóbb a kínai DeepSeek 2025 eleji piacra lépése rázta meg a piacot (CBSNews, 2025): az R1 modell a versenytársak 16 ezer GPU-s kapacitásigényével szemben állítólag mindössze 2000 GPU-t igényel a nagyon hasonló teljesítmény eléréséhez. Az R1 hirtelen jött népszerűsége nemcsak az Nvidia részvényeinek értékét csökkentette 17 százalékkal, de arra is rámutatott, hogy a piac rendkívül jól fogad minden újdonságot és erős márkahűség még nem alakult ki. Miközben több ország is vizsgálatot indított az R1 biztonsági kockázatainak felmérésére, az MI-alapú alkalmazások üzemeltetői máris megkezdték az áttérést a kínai megoldásra, mivel az általuk felszámított ár alig 0,6 százaléka az OpenAI által kérnek (Reuters, 2025).

Az MI fejlődésével párhuzamosan az informatikai cégek elkezdték azt beépíteni termékeikbe, ami az elmúlt években olyan ágazatokban hozott alapvető változásokat, mint a pénzügyek, az egészségügy vagy a gépjárműgyártás (Bughin et al., 2018). Az AI-alapú megoldások a kockázati tőke egyik vonzó célpontjává váltak, miközben napjainkra az ingyenes változatok és a hagyományosan nagy volumenben értékesített szoftverek új verzióiba bekerülő MI-asszisztensek révén az átlagfelhasználók számára is széles körben, akár ingyenesen elérhetővé vált a technológia. Csakhogy egyelőre még a legtöbb aktív felhasználónak sem világos,

milyen lényeges eltérések vannak a hagyományos és az MI-alapú IT-megoldások között. (A legfontosabb eltéréseket az *1. táblázat* foglalja össze.)

Mindez viszont azt jelenti, hogy egyre több vállalat működését érinti az MI és annak a napi használata, miközben a szervezetek nem feltétlenül vannak felkészülve azokra az új kockázatokra, amelyeket a technológia magával hoz. Ez a cikk széles körű irodalmi áttekintésre építve mutatja be, milyen kihívásokra kell felkészülni, és miképpen lehet a megjelenő új kockázatokat hatékonyan kezelni.

Először áttekintjük, milyen hatás reményében igyekeznek a cégek MI-alapú megoldásokat meghonosítani, majd a klasszikus kockázatkezelés modelljeinek alkalmazhatóságát vizsgáljuk. Azt követően rátérünk az MI-k használatának működési kockázataira és azok csoportosítására, majd a kockázatkezelési lehetőségeket vesszük sorra. Az MI-k sikeres bevezetéséhez és az új típusú kockázatkezelési módszerek meghonosításához ugyanakkor számos MI-hez kötődő paradoxont is le kell győzniük a szervezeteknek. Ezek számbavétele után az elemzés az eredmények összegzésével és javaslatok megfogalmazásával zárul.

2. AZ MI HATÁSA A VÁLLALATI MŰKÖDÉSRE

Az MI-alapú alkalmazások a vállalati működés nagyon sok területét érintik egyidőben (Brynjolfsson – McAfee, 2017). Az adatelemzési eszközök hatékonyságának növekedése például egyszerre alakítja át a kontrolling, az értékesítés és a logisztika világát, az egészségügyi diagnosztikát vagy a tőzsdei kereskedést (Bughin et al., 2018). Az MI-alapú rendszerek segítségével nem csak a működési költségek csökkenthetők, de javulhat a döntéshozatal és új innovációs lehetőségek is nyílhatnak (Davenport–Ronanki, 2018). Márpedig a költségek csökkenésével ezek az eszközök mára lényegében minden vállalkozás számára elérhetőek, vagyis aligha lesz olyan szervezet, ahol nem következnek be fundamentális változások.

1. táblázat
A hagyományos és az MI-alapú rendszerek összevetése

	Hagyományos szoftverek	MI–alapú megoldások
Felhasználási területek	Strukturált, szabályalapú környezetek (bankrendszerek, ERP, adatbázisok)	Komplex, strukturálatlan környezetek (képfelismerés, autonóm rendszerek)
Megbízhatóság	Nagyon kiszámítható, determinisztikus és szabályalapú	Kevésbé kiszámítható, valószínűségi alapon működik, az adatok minőségétől függ
Döntéshozatal	Előre meghatározott logikán és szigorú szabályokon alapul	Statisztikai modelleket és valószínűségi következtetéseket használ
Rugalmasság	Merev, a változásokhoz újraprogramozás szükséges	Alkalmazkodó, tanul az adatokból és folyamatosan fejlődik
Átláthatóság	Egyértelmű logika, könnyen auditálható	Gyakran átláthatatlan („fekete doboz" modellek), nehezen értelmezhető
Hibakezelés	A hibák egyértelműek és debuggolással javíthatók	A hibák rejtetten jelentkezhetnek, az adatoktól függenek és nehezebben visszakövethetők
Skálázhatóság	Hardver- és architektúrafüggően skálázódik	Az adatok elérhetőségétől és a számítási kapacitástól függően skálázódik
Felhasználói interakció	Strukturált bemeneteket kezel (menük, űrlapok, gombok)	Természetes nyelvet, képeket és összetett bemeneteket is feldolgoz
Alkalmazkodóképesség	Kézi frissítés és újraprogramozás szükséges	Önállóan fejlődik tanulás és finomhangolás révén
Teljesítménymérés	Pontosságon és előre meghatározott tesztelési eseteken alapul	Valós környezeti robusztusság alapján mérhető
Biztonsági kockázatok	Hackertámadások	Torzítások és modellmérgezés
Etikai megfontolások	Kevés etikai aggály, mivel előre meghatározott szabályokat követ	Etikai problémákat vet fel (elfogultság, méltányosság, elszámoltathatóság)
Adatfüggőség	Strukturált, előre meghatározott adatformátumokkal működik	Nagy mennyiségű, gyakran strukturálatlan adatból tanul
Szabályozási megfelelőség	Könnyebben szabályozható, mivel determinisztikus	Nehezebben szabályozható, mivel döntéshozatala folyamatosan változik és torzításokat tartalmazhat

Forrás: saját szerkesztés

Az AI segíthet a marketingtartalom előállításában, az üzleti döntéshozatalban és az ügyfélszolgálatban is. Egy konkrét esetben a Claude MI segítségével az átlagos panaszkezelési idő 87 százalékkal rövidült (Reuters, 2025). Az MI révén javuló adatelemző képességek sokat javíthatnak az emberi erőforrás kezelésén is (*Financial Times*, 2024). Nem csak a képzési anyagok létrehozása, a belső szervezeti képességek és tudás feltérképezése, az álláshirdetésre jelentkezők megszűrése, de a vezetőkiválasztás is gyorsabb és sikeresebb lehet. A logisztikában többek között az útvonalak optimalizálásában, a készletek tervezésében és nyilvántartásában vagy a kereslet előrejelzésében hozhatnak drámai változást az MI-alapú megoldások (Ringly, 2025).

Az egészségügyben, különösen a képalkotó diagnosztikában az MI forradalmasította a betegellátást. Például egy nemzetközi tanulmány kimutatta, hogy a mélytanulási algoritmusok pontossága a mellrák szűrésében meghaladja a radiológusok teljesítményét (McKinney et al., 2020). A pénzügyi szektorban az MI-vezérelt eszközök többek között az automatizált kockázatelemzésben, a csalások felderítésében és az algoritmikus kereskedésben nyújtanak kiemelkedő teljesítményt. Egy tanulmány szerint (Fuster et al., 2022) az MI-technológiát alkalmazó pénzügyi intézmények gyorsabb és pontosabb hitelminősítési döntéseket hoznak, csökkentve ezzel a nemteljesítési kockázatot és növelve a pénzügyi stabilitást.

Összességében úgy tűnik, hogy az MI-alapú megoldások erősen romboló jellegűek: földrengésszerű változásokat hoznak, elavulttá téve egy sor korábban használt megoldást és technológiát. (Az MI és a hagyományos IT-megoldások összevetését az *1. táblázat* mutatja.) Ráadásul a korábbi hasonló ugrásokkal ellentétben (gőzgép, gépkocsi, számítógép, internet), az MI-használóknak látszólag nem kell új képességeket megtanulniuk: az MI alkalmazkodik a felhasználóhoz, és a szövegírásban, képalkotásban vagy a forráskutatásban járatlanok is képesek igen gyorsan legalább középszerű eredményt elérni. Ez különösen azon cégeknek lehet hatalmas segítség, ahol az adott területen eddig semmilyen szakértelem sem állt rendelkezésre, vagy az adott feladatra fordítható idő nagyon csekély volt.

Az MI-technológia fejlődésével ma már nehezen különböztethetőek meg az alacsonyan képzett munkaerő és a gép által létrehozott anyagok, s több területen a közepesen képzett munkaerő munkájának minőségi előnye is eltűnni látszik. Az MI az automatizáció terjedésével veszélyeztetheti a hagyományos munkaköröket, amivel munkahelyi polarizációt és növekvő gazdasági egyenlőtlenségeket okozhat a kevésbé jó képességűek hátrányba hozásával (Hassel–Özkiziltan, 2023). Az MI igencsak szelektíven hat a munka világára: a magasan képzettek munkáját gyorsítja és egyszerűsíti, a közepesen képzettekét akár kiválthatja, miközben az alacsonyan képzettek vagy képzettség nélküliek azzal szembesülhetnek, hogy közvetlen vezetőik helyét egy MI veszi át.

E trendeket látva jogosan merül fel a legtöbb cégtulajdonosban és felsővezetőben, hogy minél előbb minél nagyobb teret kell adni a vállalati működésben az MI-t használó megoldásoknak. Csakhogy kapkodás helyett érdemes a bevezetésre alaposan felkészülni, mert máris látszik, hogy az MI használata nagyon komoly és sok tekintetben újszerű működési kockázatok megjelenésével jár.

3. FOGALMI KERETEK

Az MI okozta működési kockázatok rendszerezéséhez első lépésben érdemes áttekinteni a hagyományos kockázatkezelési keretrendszereket. Ezek közül a kockázatkezelési keretrendszer (Risk Management Framework – RMF), a három védelmi vonal (Three Lines of Defence – TLD) modellje, a COSO vállalati kockázatkezelési (ERM) keretrendszer (COSO, 2017) és a svájci sajt modell (Swiss Cheese Model – SCM) alkalmazása a legelterjedtebb. (A kockázat és biztonság informatikai modelljeiről lásd Csáki [2023] kiváló összegzését.)

Az RMF a folyamatos kockázatértékelést és monitorozást hangsúlyozza, míg a TLD a felelősségi körökre koncentrál: az operatív menedzsment (első vonal), a kockázatkezelési és megfelelőségi funkciók (második vonal), valamint a belső audit (harmadik vonal) révén biztosítja az átfogó kockázati felügyeletet. A COSO ERM-keretrendszere integrálja a kockázatkezelést a stratégiai tervezéssel, és a kockázattudatosság fontosságát hangsúlyozza a döntéshozatali folyamatokban. A svájci sajt modell ezzel szemben a kockázatok kialakulásához hozzájáruló emberi tevékenységekre fókuszál.

A hagyományos modellek azonban csak korlátozottan alkalmazhatóak az MI-vel kapcsolatos kockázatok elemzésére. A korábbi kockázati besorolásokat MI-specifikus kockázati taxonómiákra kell cserélni (MIT, 2024), a statikus kockázatértékelő modelleket folyamatos monitoringot és adaptív menedzsment megoldásokat előíró dinamikus modellekkel kell felváltani. A már használatban lévő következmény–mechanizmus–kockázat (Context-Mechanism-Risk – CMR) kockázatelemzési modellt ki kell egészíteni az MI alkalmazási környezetével, működési folyamataival és kockázataival.

Cummings (2024) azt javasolja, hogy a hagyományos SCM helyett a módszernek az MI-tevékenységre igazított, TAIHA névre keresztelt, általa kifejlesztett változatát használjuk. Ebben a fókusz az MI-alapú megoldások kialakításához és szabályozásához kapcsolódó emberi tevékenységeken van, ám a más modellekben központi szerepet kapó felhasználó–MI-interakciókat egyáltalán nem fedi le. A probléma megoldására az Egyesült Államok Kereskedelmi Minisztériuma alá tartozó National Institute of Standards and Technology 2024 nyarán egy kifeje-

zetten az MI-rendszerre szabott kockázatmenedzsment-keretrendszert publikált (NIST, 2024), amely tizenkét MI-specifikus kockázattípust azonosított.

4. MI-SPECIFIKUS MŰKÖDÉSI KOCKÁZATOK

A kockázatokat hagyományosan legalább három dimenzióban szokták csoportosítani: ok, megjelenési forma és eredet szerint. Így például egy alapanyag árának hirtelen megugrása (ok) okozhat likviditási gondokat (megjelenés), s ennek a kockázatnak azért vagyunk kitéve, mert az adott iparágban (eredet) működünk. A csoportosítás azért különösen fontos, mert sokszor vezérfonalként szolgál a kockázat kezelésekor. Ha csak arra figyelünk, hogy a kockázat melyik típusba tartozik, az árak ingadozását mint piaci kockázatot derivatív pénzügy termékek használatával mérsékelnénk, az okozott likviditási kockázatot viszont többletpénzállomány tartásával orvosolnánk, míg a kitettséget okozó iparági kockázatot tevékenységeink diverzifikálásával kezelhetjük, noha valójában egyetlen jelenség különféle vetületeit nézzük.

Az elmúlt években számos cikk és tanulmány vizsgálta az MI-hez köthető működési kockázatokat, különféle csoportosítási módokat javasolva. Például az MIT (2024) oksági taxonómiáját a *2. táblázat* mutatja míg Csáki (2023) a keskeny mesterséges intelligencia kockázatainak csoportosítására a *3. táblázat* szerinti taxonómiát javasolja. (A keskeny mesterséges intelligencia azon adatvezérelt, modellalapú intelligens rendszereket jelenti, amelyek jellemzően gépi tanulási módszerekre építve készülnek.)

2. táblázat
Az MI-hez kötődő kockázatok oksági taxonómiája

Kategória	Szint
Döntéshozó	Ember MI Egyéb
Szándék	Szándékolt Nem szándékolt Egyéb
Bekövetkezés ideje	MI-telepítés előtt MI-telepítés után Egyéb

Forrás: MIT (2024:5)

3. táblázat
A keskeny mesterséges intelligencia kockázatainak dimenzionált áttekintése

Kockázat megjelenési helye		Hiba	Gyengeség
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
	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
Szervezet		Hibásan célzott vagy rosszul feltett kérdés Nem megfelelő kontextusban történő alkalmazás Hibás feltételezések (mit jelent a szervezetnek)	Arrogáns algoritmus Átlag használó nem érti

Hatás	Hiba	Gyengeség
Egyén és szervezet viszony	Ellenőrzés elvesztése a rendszer felett Szabályozásnak nem megfelelő alkalmazás	Nem tervezett mellékhatás Autonómiával járó (felelősségi) kockázat Felelősségrevonhatóság elvesztése Befolyásolt viselkedés
Egyén és társadalom viszony	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

Forrás: Csáki (2023:44) alapján

Megint más megközelítésben a kockázatokat három fő dimenzióban vizsgálhatjuk: ezek a (I) technológiai, a (II) szervezeti és a (III) társadalmi-etikai kockázatok. A technológiai kockázatok közé tartoznak az adatvédelmi problémák, a rendszerek átláthatatlansága és a modellhibák. A szervezeti kockázatok a döntéshozatali folyamatok befolyásolását, az emberi munkaerő kiváltását és a jogi megfelelést fedik le. Végül a társadalmi és etikai kockázatok az MI által okozott diszkriminációt, a félrevezető információk terjesztését és a munkahelyi pszichológiai hatásokat foglalják magukban. Hassel és Özkiziltan (2023) szerint ugyanakkor elsődlegesen aszerint kell különbséget tennünk az MI-használat kockázatai közt, hogy azok közvetlenül vagy közvetve hatnak-e a munkavégzésre.

Az MIT (2024) ugyanakkor javasol egy megjelenési forma szerinti csoportosítást is. A következőkben ezt használva tekintjük át a szakirodalomban azonosított kockázatokat, így különösen a tizenkét osztatú NIST (2024) topológiát.

A. Diszkrimináció és toxicitás

A.1. Tisztességtelen megkülönböztetés és félrevezető ábrázolás. Az MI-alapú megoldások (jórészt a tanításukra használt dokumentumok rejtett sajátosságai miatt) hajlamosak különféle előítéletekre. Jó példa lehet erre az Amazon esete, amelynek MI-alapú kiválasztó rendszere hátrébb sorolta az IT-állásokra jelentkező nőket, mivel a korábbi évek adatai alapján a cég inkább férfiakat alkalmazott ezekben a pozíciókban (Dastin, 2018). Buolamwini és Gebru (2018) eközben arra mutatott rá, hogy az arcfelismerő rendszerek pontossága egyes rasszokban igencsak eltérő lehet, ami diszkriminatív helyzetek kialakulását idézheti elő. A NIST (2024) külön pontban említi, hogy az MI-alapú megoldásokkal a szándékosan mérgező és diszkriminatív tartalmak létrehozása is könnyebb lehet.

A.2. Káros tartalmak megosztása. Megesik, hogy az MI váratlanul toxikus tartalmaknak teszi ki a felhasználóját: előfordult már gyűlöletbeszéd, öngyilkosságra való felbujtás, illegális tevékenység támogatása vagy (gyermek) pornográfia is. A NIST taxonómiája (NIST, 2024) ezen probléma mellett önálló kockázatként említi a kémiai, biológiai, sugárzó és nukleáris (VBSN–CBRN) fegyverekkel és más veszélyes anyagokkal kapcsolatos tudás széles körben való elérhetővé tételét.

A.3. Csoportok közti egyenlőtlen teljesítmény. Az MI hajlamos arra, hogy a felhasználó különféle csoportokhoz való tartozása alapján eltérő eredményeket generáljon vagy döntéseket hozzon, főképpen a hibás rendszertervezés vagy a betanításhoz használt anyagok torzítottsága miatt. Ráadásul a szélsőséges nézeteket vallóknak az MI nagyobb eséllyel ajánlhat a nézeteiket megerősítő anyagokat, amivel a meglevő előítéleteket még inkább felerősítheti.

B. Adatvédelem és biztonság

B.1. Az adatvédelem megsértése érzékeny információk megszerzésével, kiszivárogtatásával vagy helyes kikövetkeztetésével. Az MI az érintett beleegyezése nélkül megtanulhat és másokkal megoszthat személyes információkat. Ennek egyik extrém példája a deepfake videók megjelenése. Klasszikus példa ilyen esetre a Clearview.Ai esete: az arcfelismerő technológiát fejlesztő cég több mint 3 milliárd fotót gyűjtött össze jórészt a közösségi médiából, miközben még az FBI arcfelismerő adatbázisában is csak 411 millió képet tároltak. Személyes adatok illegális gyűjtése miatt számos pert indítottak a cég ellen, hiába használta a megoldást 600 bűnüldözési szervezet is. Végül egy 2024-es bírósági döntés a károsultaknak 52 millió dolláros kártérítést ítélt meg, amivel lényegében csődbe vitte az akkor 225 millió dollárra értékelt céget. A vállalat a fizetésképtelenség elkerülésére egy későbbi tőzsdére menetelkor részvényekben ígérte kifizetni a károsultakat (Hill, 2021; 2024).

A NIST (2024) taxonómia önálló pontként veti fel a szellemi tulajdon sérülését: a különféle szerzői és szomszédos jogok által védett anyagok mindenféle engedély és jelölés nélkül bekerülhetnek az MI által generált tartalmakba. Ez alássa a társadalom jogszerű működését, és csökkenti az emberek motivációját az ilyen kreatív tartalmak előállítására, visszafogva az emberiség fejlődését.

Hassel és Özkiziltan (2023) a munkavállalók személyes adatainak minden korábbinál szélesebb körű gyűjtését és nyomon követését vetik fel. A dolgozóikat különféle érzékelőkkel ellátott RFID-kitűzők viselésére kötelező cégek adatot gyűjthetnek alkalmazottaik mozgásáról, társalgási szokásiról és szociális viszonyairól. Ráadásul egyes HR-funkciók MI-alapú alkalmazásokkal való kiváltása oda vezethet, hogy emberek életét teheti tönkre egy-egy nem kellően objektív rendszer.

Az úgynevezett „algoritmusalapú menedzsment" elmossa a magánélet és munkahely közti határokat, sértheti a személyiségi jogokat is. A cikk felvetéseit továbbgondolva, az egyénre szabott munkahelyi büntetések és jutalmak világában elképzelhető, hogy a dolgozók nem jól, hanem az MI által elvárt módon akarnak majd dolgozni. Csakhogy az MI szerint ideális munkavégzés nem feltétlenül igazodik a cég és a tulajdonosok érdekeihez, ráadásul az sem biztos, hogy mindenki éppen azonos mintázatokat követve tud a lehető legjobban dolgozni. Ezek a gondolatok már a Taylor-féle mozdulatelemzéseket is használó tudományos vezetés (scientific management) korszakát idézik (Krisztián–Nemeskéri, 2014).

B.2. Az AI-rendszerek sérülékenysége és biztonsági támadások. Ahogy minden más IT-rendszer, az MI is sérülékeny lehet informatikai támadásokkal szemben. Így elképzelhető a rendszer nem kívánt befolyásolása, vagy az abban tárolt adatok kikerülése is.

Bár ez a taxonómia külön nem említi, többek között Domokos és Sajtos (2024) is rámutat arra, hogy a hatalmas erőforrásigény miatt az MI-alapú rendszereket világszerte mindössze néhány nagy piaci szereplő működteti, ezért az MI-k szervezeti integrálása növeli a harmadikfél-kockázatot és a rendszerek sérülékenysége sem lesz már többé belsőleg menedzselhető probléma. A Boston Consulting Group felmérése szerint az MI-hibák 55 százaléka harmadik fél által készített eszközöknél jelentkezik (Cogent Infotech, 2024). Ez különösen nagy kihívás lehet a pénzügyi szolgáltató szektorban.

C. Dezinformáció és félrevezető információk

C.1. Hamis vagy félrevezető információk megosztása. Az MI hibás és megtévesztő információkat generálhat, illetve terjeszthet, ami könnyen károkat okozhat a felhasználónak. A NIST (2024) topológia emellett két külön csoportban említi a konfabulációt vagy hallucinálást (I), amikor a saját maga generálta hibás vagy téves tényeket az MI meggyőzően és határozottan valóságosnakállítja be, illetve a különféle sztereotípiákat és előítéleteket (II), azaz a rendszerszintű torzítások korábban már említett megerősítését.

C.2. Az információs ökoszisztéma szennyezése és a valóságérzékelés eltorzítása. Az MI által hibás tények alapján generált anyagok megjelenése a világhálón szennyezést okoz, miközben a felhasználó előítéleteinek megfelelő anyagok biztosítása információs buborékot alakíthat ki a felhasználó körül, aki így nem lesz képes korrigálni a nézeteit. Ehhez kapcsolódik a NIST (2024) rendszerben önálló pontot kapott információs integritás, amely arra utal, hogy a felhasználó előítéleteihez igazodva adott és torzított MI-válaszok, kikerülve a világhálóra, később keverednek a tényekkel és az emberi véleményekkel, tovább növelve a bizonytalanságot és az újabb MI-rendszerek tanuló adatbázisának torzítottságát. E kockázatra jó példa lehet a Volodimir Olekszandrovics Zelenszkijt „ábrázoló" 2022-es deepfake videó, amelyen az ukrán elnök látszólag kapitulál az oroszokkal szembeni háborúban (Pearson–Zinets, 2022).

D. Rosszhiszemű szereplők és visszaélések

Szabó (2023) rámutat: az Interpol és az Europol által kiadott anyagok szerint az MI máris megváltoztatta a bűnözést. Nemcsak a bűnelkövetés és a különösen veszélyes, valamint nagy kárt okozó anyagok receptúrájához való hozzáférés és a hatóságok félrevezetése lett könnyebb, de a deepfake-technológia miatt a büntetőeljárásárok bizonyítási folyamata is nehézkesebbé vált.

D.1. Dezinformáció, tömeges megfigyelés és befolyásolás. A külső fél által kontrollált MI-rendszerek felhasználhatók demagógiára, félrevezető kampányok lebonyolítására vagy a felhasználók megfigyelésére is.

D.2. Kibertámadások, fegyverfejlesztés vagy -használat és tömeges károkozás. Az MI-alapú rendszerek használhatóak kibertámadásokra, illetve az ahhoz szükséges eszközök kifejlesztésre.

D.3. Csalás, átverések és célzott manipuláció. Az MI-rendszerek segíthetik a különféle bűnelkövetőket, és megnövelhetik az olyan nem szándékolt bűncselekmények elkövetését is, mint például a plagizáció.

E. Ember-számítógép interakció

E.1. A felhasználó túlzott támaszkodása az MI-re valós és veszélyes MI használat. Az MI-re való túlzott anyagi vagy érzelmi támaszkodás károkat okozhat, vagy támadási pontot kínálhat másoknak. Talán a legismertebb példa erre az Egyesült Államok közlekedésbiztonsági felügyeletének a Tesla ellen indított eljárása. Egy ilyen folyamatot 2017-ben intézkedések nélkül lezártak, de az önvezető módban bekövetkezett, számos baleset nyomán 2021-ben újrakezdték vizsgálódást, ami 2 millió gépkocsi visszahívásához és szoftverfrissítéséhez vezetett. Az NHTSA 2024 októberében az újabb balesetek miatt ismét vizsgálatot indított (Shepardson–Jin, 2021; Walz, 2024). Az NIST (2024) taxonómiában külön pont az MI-k emberi vonásokkal való felruházása miatt kialakuló kockázatok, így például a pszichológiai problémák.

E.2. Az emberi önállóság és döntéshozatali képesség csökkenése. Az emberi döntések MI-vel való kiváltása a szervezetek túlzott MI-függőségéhez vezethet, miközben a döntésekből eltűnhetnek az emberi vonások és érzelmek. Az MI által felügyelt HR-döntések embertelenek lehetnek, az MI által megszervezett, tervezett és felügyelt emberi tevékenységek vezetői felelőssége tisztázatlan.

E csoportba tartozik a MIT-taxonómiából hiányzó technostresszhatás is (Ragu-Nathan et al., 2008). Például Lestari és társai (2023) kimutatták, hogy a gyorséttermekben dolgozók között az MI-vel kapcsolatos ismeretek emelik a technostresszszintet, és ez az új technológiák megjelenése miatti szorongás negatívan hat a kiszolgáló személyzet teljesítményére. Vagyis már önmagában az MI létezésének, munkahelyi megjelenési esélyének van hatása az emberek munkavégzésére. Ez a Hassel és Özkiziltan-tipológiában (2023) közvetett kockázati hatás.

F. Társadalmi-gazdasági és környezeti károk

F.1. Hatalmi koncentráció és az előnyök igazságtalan elosztása. Az MI-k felett ellenőrzést gyakorló csoportok kezében túlzottan nagy globális hatalom koncentrálódhat.

F.2. Növekvő egyenlőtlenségek és a foglalkoztatás minőségének romlása. Az MI-alapú alkalmazások elterjedése lényegesen ronthatja egyes csoportok elhelyezkedési esélyeit, miközben az MI-k működtetői anyagilag megerősödnek. Az egyesült államokbeli CFA Survey felmérése szerint a megkérdezett cégek 58 százaléka a minőség javulását, 49 százalék a kibocsátás növekedését, 47 százalékuk a munkaerőköltségek csökkenését, míg 33 százalékuk dolgozóik teljes kiváltását remélte az MI vállalati alkalmazásától (Egan, 2024).

F.3. Az emberi erőfeszítés gazdasági és kulturális leértékelése. Az MI-knek a kreatív tevékenységekben (szövegírás, programozás, grafika) játszott szerepe hátrányosan hathat az emberi kreativitásra, elértéktelenítheti az emberi erőfeszítést, és a globálisan homogén kultúra kialakulásához vezethet.

F.4. Káros versenydinamikák. Az MI-k gyors fejlődése és erős versenye a fejlesztőket arra ösztönözheti, hogy nem kellően tesztelt vagy hibás megoldásokat dobjanak piacra.

F.5. Kormányzati kudarcok és szabályozási hiányosságok. (Ezt a pontot sok csoportosítás önálló főcsoportként kezeli, nem utolsó sorban azért, mert működtetése eltérő eszközökkel és szereplőkkel valósulhat meg.) A nem megfelelő szabályozás könnyebbé teheti az MI-vel kapcsolatos visszaéléseket, és nehezebbé a kockázatkezelést. Fontos kiemelni, hogy ez a pont az eredeti taxonómia építési szempontjainak nem felel meg: itt nem az MI okozza a kockázatot, hanem az elégtelen szabályozás hat az MI-fejlesztőkre és így az MI-re. Így az ezen jelenség hatására keletkező kockázatokat a felsorolás többi pontja lefedi.

F.6. Környezeti károk. Az MI-rendszerek óriási karbonlábnyoma és energiafelhasználása hozzájárul a természeti környezet romlásához. Ebből a szempontból ugyanakkor biztató jelenség a DeepSeek megjelenése, amely versenytársaihoz hasonló teljesítményét jóval kisebb számítási igénnyel éri el.

G. Az MI-rendszerek biztonsága, hibái és korlátai

G.1 Az MI saját céljait követi, amelyek ellentétesek az emberi értékekkel és célokkal. Elképzelhető, hogy az MI hibás következtetések alapján az emberiség érdekivel ellentétesen manipulálja a felhasználókat.

G.2. Veszélyes MI-képességek kialakulása vagy kialakítása. Ha az MI-rendszerek közvetlenül képesek befolyásolni a fizikai környezetet, az nagyságrendekkel

megnöveli azon károk nagyságát, amelyeket a hibás MI-k és az MI-k rosszindulatú befolyásolása okozhat.

G.3. Képességbeli hiányosságok és megbízhatatlanság. Ha az MI-rendszerek bizonyos körülmények között megbízhatatlanná válnak, az súlyos károkat okozhat egyes kritikus rendszerekben. Az Egyesült Államok közlekedési felügyelete, a National Transportation Safety Board egy önvezető Uber-gépkocsi balesete kapcsán a kiváltó okot az elégtelen tesztelésben és a megfelelő biztonsági eszközök hiányában azonosította (NTSB, 2019). Hasonló módon a – hibás – nagy gyakoriságú kereskedési (HFT) algoritmusokat szokás okolni a pénzügyi piacokon újabban feltűnt „flash crash" jelenségekért, amikor egy termék árfolyama másodpercek vagy percek alatt óriásit esik, majd visszakapaszkodik az eredeti szintre (Majumder–Yashraj, 2024). Az emberekkel együtt dolgozó, kollaboratív robotok (cobotok) baleseteket és emberi sérülést okozhatnak, az MI-alapú orvosi berendezések hibája pedig téves diagnózishoz és kezeléshez vezethet.

G.4. Átláthatóság vagy értelmezhetőség hiánya. Az MI-k döntéseinek átláthatatlansága gyengíti a döntésekbe vetett bizalmat, nehezíti a hibák kijavítását és a felelősök számonkérését. Ez különösen a kritikus rendszerek irányításánál, a gyógyászatban és pénzügyi alkalmazásoknál (Domokos–Sajtos, 2024) lehet súlyos gond.

G.5. Az MI jólléte és jogai. Az MI képességeinek fejlődésével egyre inkább előtérbe kerül a rendszerek jogalannyá válása és jólléti minimumaik meghatározása, ami további etikai kérdéseket nyit meg, és újfajta kockázatok megjelenéséhez vezethet. Már tudományosan is igazolt (Yin et al., 2024), hogy a kedvesebben megfogalmazott kéréseket (promptokat) jobb minőségben válaszolják meg az MI-k, de az optimális kedvességi szint nyelvenként eltérő. Nem tudhatjuk, hogy hajlamosabb-e egy MI hallucinálni vagy téves adatokat adni a durvább kérdésekre, esetleg egy MI-alapú rendszer képes lehet-e bizonyos felhasználókkal szimpatizálni, másokat pedig negatívan megkülönböztetni.

Fontos látni, hogy az MIT-taxonómia (2024) a megjelenési formára, kockázati forrásra koncentrál, miközben sok hagyományos besorolás a károk jellegére fókuszál. Így például az előbbiek szinte mindegyike okozhat reputációs kockázatot (Holweg et al., 2022), ha cég megítélése sérül egy-egy incidens miatt. Ugyancsak szembesülhetünk likviditási kockázattal, ha egy hiba jelentős anyagi kárt okoz, vagy szabályozási kockázattal, ha kiderül, hogy egy rendszerünk nem az előírások szerint üzemel.

5. KOCKÁZATKEZELÉSI MEGOLDÁSOK

Adatvédelem. Robosztus adatvédelemre a MI-alkalmazások mindkét oldalán szükség van. Egyrészt a torzítások, kibertámadások és a szerzői jog megsértésének elkerülésére a tanuló adatbázisokat kell alaposan megszűrni, másrészt a kimenti oldalon kiadható információ jellege is függhet a felhasználó személyétől és a felhasználási környezettől (NIST, 2024).

Hozzáférés-szabályozás. Korlátozni kell az egyes MI-alapú rendszerekhez való hozzáférést, hogy a különféle támadások esélyét csökkentsük (NIST, 2024).

Változatos és reprezentatív tanuló adatkészletek alkalmazása. Ahelyett, hogy az MI tanítását válogatás nélkül mindenféle forrásból összeollózott anyagokra bíznánk, jól megválogatott, a valóságot helyesen bemutató adatokkal érdemes dolgozni, elfogultság-felismerő algoritmusokat és etikai felügyeletet kell bevezetni.

Kiterjedt tesztelés és folyamatos monitoring. Úgy tűnik, a mai MI-rendszerek nem alkalmazhatók folyamatos teljesítményfigyelés nélkül, mert csak így vehető észre, ha az előzetesen elvégzett stressztesztek nem derítettek fel valamilyen hibát, vagy az MI és a valóság változása miatt új hiba keletkezett, esetleg az alkalmazott hibatűrő mechanizmusok nem tökéletesek. Ez a tesztelés nem csak az MI-ket gyártó és használó cégek feladata lehet: a kormányzatoknak és szabályozó hatóságoknak is végezniük kell ilyen feladatot, mert az interneten elérhető, ingyenes MI-megoldásokat olyan szervezetekben is használni fogják, amelyek esetleg az ellenőrzéséhez szükséges szaktudással nem rendelkeznek. Ilyenek lehetnek például Magyarországon a mikro- és kisvállalkozások. Ezek a szereplők a beszállítói láncon keresztül és a foglalkoztatásban játszott szerepük miatt komoly gazdasági kockázatot okozhatnak, ha az ingyenesen elérhető MI-k valamilyen súlyos hibájára nem derül időben fény.

Szabályozás. Ahhoz, hogy az MI fejlesztése kellő figyelemmel és társadalmilag is helyes célokat követve történjen, megfelelő jogszabályokra, etikai irányelvekre és a számonkérhetőség fenntartására van szükség. Az MI-kkel kapcsolatos általános elvárásokat épp úgy közzé kell tenni, mint például a gépjárműveknél, hogy az MI-k gyártói be tudják építeni a társadalmi igényeket a fejlesztési folyamataikba és a tesztjeikbe. Ebben a tekintetben különösen fontos a határokon átívelő, akár globális szabályozói együttműködés, hiszen az interneten keresztül határokra tekintet nélkül elérhetőek lehetnek a különféle, változatos minőségű MI-alapú megoldások.

Mindezek tetejébe az MI-alkalmazások – ellentétben a hagyományos informatikai rendszerekkel – folyamatosan tanulnak és fejlődnek. Ez előre nem látható viselkedéshez és nem szándékolt következményekhez vezethet, ha nem megfelelően felügyelik és kontrollálják ezeket a rendszereket. Így biztosan nem elégséges

egyszer megfelelő szabályozási környezetet kialakítani, azt folyamatosan a változásokhoz kell majd igazítani (Európai Bizottság, 2025).

Ráadásul az MI-vel kapcsolatos működési kockázatok sokrétűek, sokszor messze túlmutatva a vállalati működés keretein. Magukban foglalják ugyanis a technikai hibákon túl a stratégiai, reputációs, jogi, egészségügyi, pszichológiai és társadalmi-gazdasági kihívásokat is. Ezek a kockázatok zavarokat okozhatnak az üzleti folyamatokban, aláásható a érdekelt felek bizalmát, és jelentős pénzügyi, illetve jogi következményekkel járhatnak (Cummings, 2024).

A dolgozók oktatása. A munkáltatóknak folyamatos képzéssel növelniük kell az alkalmazottak MI-vel kapcsolatos ismereteit, hogy a szakszerűtlen használat miatti kockázatokat mérsékeljék. Annak érdekében is szükség lehet képzésekre, hogy a feladatok eddigi elvégzői képesek legyenek az immár MI-k által generált végeredmények elfogadhatóságának megítélésére, a tények ellenőrzésére és a hallucinációk kiszűrésére.

A munkatársak pszichológiai felkészítése. A szervezeteknek külön figyelniük kell arra, hogy az MI-alapú rendszerek megjelenési lehetősége vagy az azokkal végzett napi munka milyen pszichikai kihívásokat jelenthet a dolgozóknak és támogató munkahelyi környezet kialakításával csökkenteniük kell a negatív hatásokat. Hosszabb távon érdemes felkészülni arra is, amikor 15-20 év múlva a munkaerőpiacon megjelenik majd az a generáció, amelynek tagjai bennszülöttként használják majd az MI-eszközöket, miközben a hagyományos, MI nélküli módszereket nem feltétlenül képesek alkalmazni.

4. táblázat
A mesterséges intelligencia alkalmazásának kockázatai és kezelési módjai

Kockázati kategória	Kockázattípus	Kockázat leírása	Legjobb kezelési módszerek
(I) Technológiai kockázatok	Diszkrimináció és toxicitás	Az MI döntéshozatalában rejlő előítéletek és szisztematikus torzítások	Reprezentatív és elfogulatlan adatkészletek használata, etikai felügyelet
	Adatvédelem és biztonság	Személyes adatok kiszivárgása, adatlopás és jogi megfelelési problémák	Erős adatvédelmi intézkedések, hozzáférésszabályozás, titkosítás
	Rosszhiszemű szereplők és visszaélések	Az MI-rendszerek felhasználása kibertámadásokra, csalásokra és visszaélésekre	Kiberbiztonsági intézkedések, csalásmegelőző technológiák bevezetése
	MI-rendszerek biztonsága és hibái	A rendszerek megbízhatatlansága, átláthatóság hiánya és biztonsági kockázatok	Folyamatos tesztelés, átláthatósági mechanizmusok, szabványok betartása
(II) Szervezeti kockázatok	Dezinformáció és félrevezető információk	Az MI által generált téves vagy félrevezető információk elterjedése	Folyamatos monitoring, megbízható források integrációja, felhasználói oktatás
	Ember-számítógép interakció	Az emberi önállóság csökkenése, túlzott MI-függőség és technostressz	Munkahelyi oktatás, etikai és pszichológiai támogatás, emberi felügyelet
(III) Társadalmi-etikai kockázatok	Stratégiai és szabályozási kihívások	Szabályozási hiányosságok, felelősségi kérdések és stratégiai kihívások	Nemzetközi szabályozások összehangolása, felelősségi körök meghatározása
	Társadalmi-gazdasági és környezeti károk	Az MI munkahelyi és társadalmi hatásai, egyenlőtlenségek növekedése	Munkahelyi alkalmazkodási stratégiák, társadalmi felelősségvállalás

Forrás: saját szerkesztés

Szervezeti és etikai szabályzatok kiegészítése. A ma már sok helyen elérhető etikai szabályzatokat, valamint a belső szervezeti és ügymeneti utasításokat célszerű úgy frissíteni, hogy azok az MI-kkel kapcsolatos szabályokat is tartalmazzák, még akkor is, ha a cég maga nem alkalmaz MI-alapú megoldásokat. A tapasztalat ugyanis azt mutatja, hogy az alkalmazottak egy része saját munkájának megkönnyítésére kíváncsiságból és önszorgalomból is elkezdi az ingyenes MI-k használatát, s megfelelő szabályozás és képzés hiányában ez komoly kockázatokat rejt.

Incidenskezelési tervek. Mint minden kockázati típusnál, az MI-khez kötődőknél is célszerű válságterveket kidolgozni és ehhez kapcsolódó oktatást szervezni, hogy a szervezet felkészült legyen arra, ha az MI-k hirtelen használhatatlanná válnának valamilyen hiba vagy külső támadás miatt. Éppen ezért egyetlen vál-

lalkozás sem engedheti meg magának, hogy az MI-kkel kiváltott tevékenységhez szükséges emberi szakértelmet teljesen leépítse a szervezeten belül. Inkább az volna a célszerű, ha a kiváltott alacsony és közepes szaktudás helyett kisebb létszámú, de magas szintű szaktudást birtokló alkalmazottal dolgoznának. Ezzel képesek lennének az MI munkájának ellenőrzésére, illetve felkészülnének az MI-k esetleges kiesésére is.

A különféle kockázati kategóriákhoz tartozó legjobb kockázatkezelési gyakorlatokat a szakirodalom alapján a *4. táblázat* mutatja. Ugyanakkor e gyakorlatok bevezetését többféle paradoxon is nehezíti.

6. MESTERSÉGES INTELLIGENCIA-PARADOXONOK

Az MI-alkalmazásokkal kapcsolatos kockázatkezelést lényegesen bonyolultabbá teszik a szakirodalomban AI-paradoxonokként aposztrofált jelenségek. Ezek az MI fejlesztésének és alkalmazásának ellentmondásos területei. Ezek hátteréről kitűnő összegzést kínál az e fejezet elsődleges forrását adó, hét paradoxont bemutató Bakonyi (2024). Hasonló kutatás alapján Jazairy és társai (2024) a vállalati tervezésben összesen tizenkét paradoxont azonosítottak, míg más szerzők egy-egy ellentmondásra koncentráltak cikkeikben.

1.) *MI-stabilitási paradoxon* (AI Stability Paradox). Ez a jelenség azt írja le, hogy az MI-rendszereket, és különösen a neurális hálókat eredetileg a különféle folyamatok nagyobb pontosságú, stabil leírására fejlesztették ki, miközben tudjuk, hogy egyes problémákra nem lehet ilyen rendszert építeni (University of Cambridge, 2022), különösen, ha a leírni kívánt jelenség vagy probléma időben maga is változik.

2.) *Generatív MI-paradoxon* (Generative AI Paradox). A paradoxon azt írja le, hogy az MI-k képesek lehetnek akár szakértői szintű (vagy annak tűnő) tartalmak előállítására is, miközben semmilyen valódi tudásuk vagy megértésük nincsen az adott jelenségről. Így a tartalom úgy is meggyőző lehet, ha pontatlan és hibás. Ez azért veszélyes, mert az MI lényegében csak úgy tesz, mintha szakértő lenne, és a hozzá nem értőket ez félrevezetheti, különösen, mert (a) az embereknél a megértés megelőzi a szakértői szintű anyagok elkészítésének képességét, és (b) a hagyományos informatikai eszközöknél éppen a pontosság és a részletekre és tényekre támaszkodás a megszokott (West–Aydin, 2024). Ráadásul (c) az embereknél az MI-kéhez hasonló magabiztosság inkább a mélyebb összefüggéseket is átlátókra jellemző.

3.) *MI bizalmi paradoxon* (AI Trust Paradox). A paradoxon azt emeli ki, hogy a technológia elfogadása és az abba vetett bizalom korántsem jár együtt.

A megfigyelések szerint sokan úgy használják az MI-alapú rendszereket, hogy közben (helyesen) nem bíznak azokban. Így viszont az MI segítségével előállított tartalmakban akkor is kételkednek, ha erre egyébként nem lenne okuk, mert az emberi szerzők minden tényt alaposan ellenőriztek és kellő hozzáértéssel rendelkeznek.

4.) *Szakértői paradoxon* (Domain expert paradox). Ez az ellentmondás szorosan kapcsolódik a bizalmi és a generatív paradoxonhoz. Azt a jelenséget írja le, amely szerint az emberek jobban bíznak az olyan algoritmusokban, amelyek kifejlesztésében szakértők is részt vettek. Csakhogy a szakértőknek nem érdeke a közreműködés, mert végső soron saját maguk kiváltásán dolgoznak (Jazairy et al., 2024). Ráadásul szakértelmüknek köszönhetően elsősorban éppen ezek a szakemberek veszik észre az MI-k hibáit, így ők bíznak azokban a legkevésbé. Miközben tehát nekik fűződik a legkevesebb érdekük a hibák kijavításához, ők a legalkalmasabb a javítások elvégzésére.

5.) *Tudáshelyettesítési paradoxon* (Knowledge substitution paradox). Az MI képes bizonyos szintű szervezeti tudást egyes területeken kiváltani, azonban ahhoz, hogy az MI által generált eredményt szakmailag ellenőrizni tudjuk, a kiváltottnál magasabb szintű tudásra lehet szükség. Ráadásul valós logikai megértés és összefüggés-ismeret hiányában az MI csak a múltbeli adatokra tud támaszkodni, új helyzetekre aligha reagál jól. Emiatt az olyan szervezetekben, ahol a minőség szempontjából kritikus területen akarják az MI-alapú megoldásokat használni, nem feltétlenül keletkezik nagyobb megtakarítás. Különösen igaz ez, ha az adott feladatot az MI-k esetleges kiesésekor is el kell tudni végezni. Jazairy és társai (2024) rámutatnak: a szervezetnek MI-alapú eszközök bevezetésekor döntenie kell arról is, hogyan lesz elérhető és felhasználható a továbbiakban a korábban az alkalmazottak fejében és a vállalati tudástárban felhalmozott szakterületi ismeret. Fontos az is, hogy egy külső partner által működtetett MI csak nagyon korlátozottan rendelkezhet olyan cégspecifikus tudással, mint például a partnerek vagy az alkalmazottak adatai vagy a belső szabályzatok.

6.) *Kreativitásparadoxon* (Creativity Paradox). Az MI-alapú eszközök használata a gyengén vagy közepesen képezett munkavállalók kreativitását lényegesen javíthatja, ami nagy segítség lehet, ha az adott szaktudás ennél magasabb szinten nem is állt soha rendelkezésre a szervezetben. Ha azonban minden konkurens ilyen eszközöket kezd el használni, az egyediség eltűnik és tömegszerűségbe fordul, hiszen a „kreatív" anyagok nem újak, csak a múltban mások által készített anyagok iterálásai (Osadchaya et al., 2024). Ekkor már ismét csak a magas szintű, valódi emberi tudás lehet majd az üzleti siker kulcsa.

7.) *Feladatkiváltási paradoxon* (Task substitution paradox). Az MI-ket elvileg az alkalmazottak munkájának könnyítésére és gyorsítására alkalmazzák, de általánosak azok a várakozások, hogy a felszabadult munkaidőt a munkáltatók a létszám leépítésére használják. Így a megmaradó kollégák nem dolgoznak majd kevesebbet, csak másképpen, a megtakarítást pedig nem a munkakörülményeket javítja, hanem a vállalati hozamokat. Sőt előfordulhat, hogy a magasabb hozamokat az eladásoldali árverseny felemészti, ezért azok, akik megőrzik a munkájukat, olcsóbban vásárolhatnak, de a közben felszabaduló munkaerő lenyomja a béreket is. Így végső soron senki sem jár jól az MI-k alkalmazásával. Más érvek szerint (Ferraro et al., 2024) az MI teremt is új munkahelyeket, vagyis a technológia egyszerre rombol és alkot.

8.) *Időparadoxon* (Time paradox). Miközben az MI-k a különféle feladatok elvégzéséhez szükséges idő rövidülését ígérik, a mesterségesintelligencia-alapú megoldások bevezetése, finomhangolása, betanítása a hagyományos IT-rendszerekhez hasonlóan igen időigényes és – szemben a klasszikus IT-rendszerekkel – később az MI működésének és eredményeinek állandó ellenőrzése folyamatos többletfeladatként jelentkezik (Osadchaya et al., 2024). Így igazi megtakarítást csak közép- és hosszú távon remélhetünk inkább.

9.) *Hibaparadoxon* (Error paradox). A hagyományos számítógépes megoldások jellemzően pontosabbak az emberi munkánál, s a jól implementált MI-megoldások még tovább javíthatják a pontosságot. Csakhogy míg az emberi hibák kisebbek és gyakoribbak, az MI ritkábban, de jóval nagyobbakat hajlamos hibázni. Így a kockázatkezelésnek a gyakori, de kis hatású események helyett ritka, de nagyobb hatású problémákra kell felkészülnie, ami sokszor jóval nehezebb. Ráadásul ezeket a hibákat a szervezet is másképpen éli meg: az emberek ugyanis hajlamosak sokkal elnézőbbek lenni egymás hibái iránt, mint az IT-rendszerekkel szemben. Scaringi és társai (2024) például egy olyan esetet mutatnak be, ahol a nagyérelzáródás (large vessel occlusion – LVO) diagnosztizálására kidolgozott, egyébként nagyon jól teljesítő algoritmust azért ítélték egy klinikán megbízhatatlannak az orvosok, mert egy korábbi esetben fals pozitív jelzést adott.

10.) *Referenciaparadoxon* (Reference paradox). Az emberek hajlamosabbak hinni az olyan előrejelzéseknek és értékeléseknek, amely csak kis mértékben tér el saját várakozásaiktól. Ez viszont oda vezet, hogy az MI-alkalmazások paraméterezésénél az emberi szakértők várakozásaihoz hasonló eredményt adó algoritmusokat ítélik jobbnak. Ha azonban az MI eredményei nagyon hasonlóak, kétségek merülhetnek fel a rendszer hasznosságát illetően.

11.) *Tapasztalati paradoxon* (Experience paradox). Az MI-megoldások a tanulási adatbázis mintázatainak felismerésében jók, s kizárólag számok alapján mérlegelnek. Ugyanakkor az emberi szakértők nagyon sokszor kvalitatív

szempontokat is figyelembe vesznek, és a nehezen átlátható, komplex matematikai modellek helyett inkább hisznek a hétköznapi, verbálisan könnyebben leírható tapasztalatokban és a megérzésekben (Jazairy et al., 2024).

A bizalmatlanságot az MI által feltárt meglepő számszaki összefüggések bemutatásával lehet mérsékelni, de ha ez ellentmond az emberi tapasztalatnak, akkor könnyen felmerül a gyanú, hogy csak valamilyen múltbeli egyszeri esemény torzította a mintát. Az MI logikájának széles körű elfogadásához sok olyan tapasztalatra van szüksége a szakértőknek, amikor utóbb a gépi eredmények bizonyultak helyesnek. Ez azonban időigényes, és bekövetkezése csak akkor reális, ha a referenciaparadoxont úgy kerüli meg a szervezet, hogy az MI-t nem a korábbi szakértők helyett, hanem azokkal párhuzamosan alkalmazza.

12.) *MI-illeszkedési paradoxon* (AI Alignment Paradox). Ez a jelenség arra utal, hogy minél inkább sikerül az MI-t az emberi gondolkodáshoz és értékrendszerhez közelíteni, annál sebezhetőbbé válik a rosszindulatú befolyásolással szemben (West–Aydin, 2024). Ez pedig nem cél, vagyis hiba volna az MI-t „teljesen emberivé" tenni, noha az jelentősen javíthatná elfogadottságát.

13.) *Felsőbbrendűségi paradoxon* (Superiority Paradox). Az MI-vel dolgozó emberek egyidejűleg érzik magukat alsóbbrendűnek és felsőbbrendűnek is (Osadchaya et al., 2024). Miközben elmaradnak tárgyi tudásban, sebességben és gyakrabban hibáznak, messze túlszárnyalhatják azt kreativitásban, az összefüggések megértésében és reálisabb önértékelésükkel.

14.) *Hamis kapcsolat paradoxon* (Illusive connection paradox). A valós személyeket utánzó AI-chatbotok az ügyfelekben a személyes kapcsolat illúzióját keltik. Ha azonban kiderül, hogy az MI-t tévesen gondolták embernek, az visszaüthet, és éppen ellenkező hatást válthat ki (Ferraro et al., 2024).

15.) *Elégedettségi paradoxon* (Satisfaction paradox). Az MI-alapú ügyfélszolgálatok gyorsabban dolgoznak és pontosabb információt adhatnak, ami növelheti az ügyfél-elégedettséget. Ugyanakkor az empátiát igénylő vagy egyedi esetekben az érintettek még elégedetlenebbek lehetnek, ha egy gép képtelen segíteni rajtuk, élő embert azonban nem tudnak elérni (Ferraro et al., 2024).

Jazairy és társai (2024) ezeken túl néhány olyan problémát is bemutatnak, amelyek inkább dilemmák, mint paradoxonok. Ilyen dilemma az, hogy vajon azzal jár-e egy cég jobban, ha a gyorsabb és átlátható hagyományos tervezési eszközöket használja, és megvárja a fejlettebb MI-eszközök elkészültét, vagy ha úttörőként adaptálva az esetleg kevésbé pontos rendszereket, piaci előnyre tehet szert.

Tisztázni kell azt is, hogy a tervezésben reaktívak vagy proaktívak igyekszünk lenni, vagyis a tűzoltásszerű beavatkozások mennyire reálisak és költségesek, az a

tervezési hibák milyen károkat okoznak. Az MI bevezetésénél dönteni kell centralizált és decentralizált rendszerek között, a lépésenkénti vagy egyidejű bevezetésről, és arról is, hogy mennyire vagyunk hajlandók áttervezni meglevő folyamatainkat a MI-hez igazodva.

A HR-stratégiában döntést kell hozni arról, hogy a jövőben MI-ismeretekkel is rendelkező hagyományos szakértőket akarunk alkalmazni, vagy elsősorban általános MI-gurukat (Jazairy et al., 2024). A legnagyobb dilemma azonban talán mégis az, hogy az MI-eszközök használatát akarjuk integrálni a meglevő szervezeti keretekbe vagy a teljes működést és a munkaköröket szervezzük az MI köré.

A feltárt paradoxonok rámutatnak arra, hogy az MI vállalati bevezetése nem csupán technológiai, hanem stratégiai és filozófiai kihívás is. Az MI egyszerre növelheti a hatékonyságot és generálhat új problémákat, miközben az emberi munkát kiegészítheti, vagy teljesen kiválthatja. E paradoxonok ismerete elengedhetetlen a megalapozott kockázatkezelési stratégiák kidolgozásához. A vállalatok számára ez azt jelenti, hogy az MI-t nem csupán eszközként kell kezelniük, hanem egy olyan tényezőként, amely alapjaiban formálja át a működési környezetet, ezért annak folyamatos felügyelete és adaptív szabályozása szükséges.

7. LEGFONTOSABB EREDMÉNYEK

Összességében elmondható, hogy az MI-alapú megoldások bevezetése és működtetése egészen más jelleghű kihívásokat tartogat, mint amit a hagyományos IT-megoldásoknál megismerhettünk. Az MI-alapú rendszerek valós megértés nélkül képesek szakértőkhöz hasonló minőségű anyagokat készíteni, s magabiztosságukkal könnyedén megtéveszthetik a hagyományos IT-rendszerek pontosságához szokott felhasználót. Emiatt folyamatos felügyeletet igényelnek, és eredményeiket részletekbe menően ellenőrizni kell. Ráadásul használatuk sok esetben csak középszintű szakismereti szintig képes igazán javítani a hatékonyságon, afelett a kreativitás eltűnésével tömegszerű eredményeket kaphatunk.

Mivel az MI-alapú megoldások egészen új támadási felületeket is jelentenek, működési módjuk pedig a hagyományos eszközökkel csak nehezen és korlátozottan érthető meg, nagyon fontos, hogy a piaci változásokhoz folyamatosan alkalmazkodó etikai és jogi normák irányítsák a fejlesztéseket, világos felelősségi körökkel és a már működő rendszerek állandó monitorozása mellett.

Bár a mesterséges intelligencia (MI) a modern munkahelyek szerves részévé vált, és sok esetben jelentős átalakulást hozott az üzleti hatékonyságban, innovációban és döntéshozatalban, a technológia még gyerekcipőben jár. Alkalmazása igen összetett működési kockázatokkal jár, amelyek technikai, etikai, jogi és társadalmi-gazdasági dimenziókban egyaránt megjelennek. E kockázatok csökkentéséhez

elengedhetetlen az MI-vel kapcsolatos ismeretek terjesztése még az olyan szervezetekben is, ahol hivatalosan nem is működtetnek ilyen rendszereket.

Az MI-technológiák fejlődésével és elterjedésével egyre fontosabbá válik a robusztus, etikus és adaptív kockázatkezelés. Azok a szervezetek, amelyek proaktívan kezelik az MI-vel kapcsolatos kockázatokat, folyamatosan tanulnak, és elkötelezettek az etikus irányítás mellett, versenyelőnyre tehetnek szert, miközben megóvják működésük fenntarthatóságát és társadalmi felelősségvállalásukat.

Az MI működési kockázatainak kezelése nem csupán technikai vagy szabályozási kihívás, hanem stratégiai feladat is. A vállalatok számára elengedhetetlen az adaptív kockázatkezelési modellek bevezetése, az MI-specifikus etikai szabályok kidolgozása és a folyamatos monitoring. A szabályozó hatóságoknak pedig olyan kereteket kell létrehozniuk, amelyek egyszerre ösztönzik az innovációt, és minimalizálják a visszaélések lehetőségét.

Összegezve: 1) a vállalatoknak olyan átláthatósági mechanizmusokat kell bevezetniük, amelyek révén az MI döntései ellenőrizhetők és visszakövethetők lesznek. 2) Proaktív kockázatkezelési rendszereket kell kialakítaniuk, amelyek lehetővé teszik az MI működésének folyamatos értékelését. Végül 3) munkavállalói képzések kell indítaniuk és etikai szabályozásokat kell elfogadniuk, hogy az alkalmazottak tudatosan és biztonságosan használhassák az MI-alapú eszközöket.

A szabályozók feladata olyan a) egységes jogi keret létrehozása, amely világosan meghatározzák az MI-fejlesztők és felhasználók felelősségét. Ehhez elengedhetetlen a b) nemzetközi együttműködés, hiszen az MI-k hatása globális, így a kiskapuk elkerülése érdekében a szabályozásuk is összehangolt kell legyen. Végül c) hatékony felügyeleti mechanizmusokat kell kiépíteni, amelyek biztosítják az MI-alapú rendszerek állandó felügyeleti monitoringját.

HIVATKOZÁSOK

AI Now Institute (2019): *AI Now Report 2019*. https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2019_Report.pdf (letöltve: 2025.02.07.).

Bakonyi, Z. (2024): How can companies handle paradoxes to enhance trust in artificial intelligence solutions? A qualitative research. *Journal of Organizational Change Management*, 37(7), 1405–1426. https://doi.org/10.1108/JOCM-01-2023-0026.

Brynjolfsson, E. – McAfee, A. (2017): The business of artificial intelligence. *Harvard Business Review*. https://hbr.org/2017/07/the-business-of-artificial-intelligence (letöltve: 2025.02.07.).

Bughin, J. – Seong, J. – Manyika, J. – Chui, M. – Joshi, R. (2018): Notes from the AI frontier: Modeling the impact of AI on the world economy. *McKinsey Global Institute*. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy (letöltve: 2025.02.06.).

Buolamwini, J. – Gebru, T. (2018): Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 77–91.

CBSNews (2025): What is DeepSeek, and why is it causing Nvidia and other stocks to slump? https://www.cbsnews.com/news/what-is-deepseek-ai-china-stock-nvidia-nvda-asml/ (letöltve: 2025.02.06.).

Cogent Infotech (2024): AI Risks: How Businesses Can Safeguard Their Future. https://www.cogentinfo.com/resources/ai-risks-how-businesses-can-safeguard-their-future (letöltve: 2025.02.09.).

COSO (2017): Enterprise Risk Management: Integrating with Strategy and Performance. https://www.coso.org/_files/ugd/3059fc_61ea5985b03c4293960642fdce408eaa.pdf (letöltve: 2025.02.07.).

Cummings, M. L. (2024): A Taxonomy for AI Hazard Analysis. *Journal of Cognitive Engineering and Decision Making*, 18(4), 327–332. https://doi.org/10.1177/15553434231224096.

Csáki Cs. (2023): A mesterséges intelligencia elterjedéséből adódó kockázatok szisztematikus vizsgálata. In: Kovács, Z. (ed., 2023): *A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata. Budapest, Katonai Nemzetbiztonsági Szolgálat*, 27–50.

Dastin, J. (2018): Insight – Amazon scraps secret AI recruiting tool that showed bias against women. https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (letöltve: 2025.02.08.).

Davenport, T. – Ronanki, R. (2018): Artificial Intelligence for the Real World. *Harvard Business Review,* 96(1), 108–116. https://hbr.org/2018/01/artificial-intelligence-for-the-real-world.

Devlin, J. – Chang, M. W. – Lee, K. – Toutanova, K. (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:*1810.04805. https://arxiv.org/abs/1810.04805.

Domokos, A. – Sajtos, P. (2024): Mesterséges intelligencia a pénzügyi szektorban – Innováció és kockázatok. *Hitelintézeti Szemle*, 23(1), 155–166.

Egan, M. (2024): AI is replacing human tasks faster than you think. https://edition.cnn.com/2024/06/20/business/ai-jobs-workers-replacing/index.html (letöltve: 2025.02.09.).

Európai Bizottság (2025): A mesterséges intelligenciáról szóló rendelet. https://www.consilium.europa.eu/hu/policies/artificial-intelligence/ (letöltve: 2025.02.07.).

Ferraro, C. – Demsar, V. – Sands, S. – Restrepo, M. – Campbell, C. (2024): The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons*, 67(5), 549–559. https://doi.org/10.1016/j.bushor.2024.04.013.

Financial Times (2024): Employers look to AI tools to plug skills gap and retain staff. https://www.ft.com/content/9cf58a76-5245-4cdf-9449-239e90077eb5 (letöltve: 2025.02.07.).

Hassel, A. – Özkiziltan, D. (2023): Governing the work-related risks of AI: implications for the German government and trade unions. *Transfer: European Review of Labour and Research*, 29(1), 71–86. https://doi.org/10.1177/10242589221147228.

Hill, K. (2021): The Secretive Company That Might End Privacy as We Know It. *The New York Times*. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (letöltve: 2025.02.09.).

Hill, K. (2024): Clearview AI Used Your Face. Now You May Get a Stake in the Company. *The New York Times*. https://www.nytimes.com/2024/06/13/business/clearview-ai-facial-recognition-settlement.html (letöltve: 2025.02.09.).

Holweg, M. – Younger, R. – Wen, Y. (2022): The reputational risks of AI. *California Management Review*. https://cmr.berkeley.edu/2022/01/the-reputational-risks-of-ai/ (letöltve: 2025.02.08.).

Jazairy, A. – Shurrab, H. – Chedid, F. (2024): Impact pathways: walking a tightrope—unveiling the paradoxes of adopting artificial intelligence (AI) in sales and operations planning. *International Journal of Operations & Production Management*, 45(13), 1–27. https://doi.org/10.1108/IJOPM-07-2024-0582.

Kreps, S. – George, J. – Lushenko, P. – Rao, A. (2023): Exploring the Artificial Intelligence ‚Trust Paradox': Evidence from a Survey Experiment in the United States. *PLOS ONE*, 18(1), e0288109. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0288109 (letöltve: 2025.02.09.).

Krisztián, B. – Nemeskéri, Z. (2014): A Taylori elvek a magyar gazdaságban. *Taylor Gazdálkodás- és Szervezéstudományi Folyóirat*, 6(1-2), 498–508., https://ojs.bibl.u-szeged.hu/index.php/taylor/article/view/12838 (letöltve: 2025.02.19.).

Krizhevsky, A. – Sutskever, I. – Hinton, G. (2012): ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25. https://doi.org/10.1145/3065386.

Lestari, N. S. – Rosman, D. – Veithzal, A. P. – Zainal, V. R. – Triana, I. (2023): Analysing the Impact of Robot, Artificial Intelligence, and Service Automation Awareness, Technostress and Technology Anxiety on Employees's Job Performance in The Foodservice Industry, *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, Pangkalpinang, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICORIS60118.2023.10352286.

Majumder, S. – Yashraj, A. (2024): Mitigating AI-Driven Flash Crashes. http://dx.doi.org/10.2139/ssrn.4950688 (letöltve: 2025.02.08.).

McKinney, S. M. – Sieniek, M. – Godbole, V., et al. (2020): International evaluation of an AI system for breast cancer screening. *Nature, 577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6.

MIT (2024): The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence, https://arxiv.org/pdf/2408.12622 (letöltve: 2025.02.07.).

NIST (2024): Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, https://doi.org/10.6028/NIST.AI.600-1 (letöltve: 2025.02.07.).

NTSB (2019): Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018, Accident Report https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf (letöltve: 2025.02.08.).

Osadchaya, E., Marder, B. – Yule, J. A. – Yau, A. – Lavertu, L. – Stylos, N. – Oliver, S. – Angell, R. – Regt, A. de – Gao, L. – Qi, K. – Zhang, W. Z. – Zhang, Y. – Li, J. – AlRabiah, S. (2024): To ChatGPT, or not to ChatGPT: Navigating the paradoxes of generative AI in the advertising industry. *Business Horizons*, *67*(5), 571–581. https://doi.org/10.1016/j.bushor.2024.05.002.

Pearson, J. – Zinets, N. (2022): Deepfake footage purports to show Ukrainian president capitulating. https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16/ (letöltve: 2025.02.09.).

Ragu-Nathan, T. S. – Tarafdar, M. – Ragu-Nathan, B. S. – Tu, Q. (2008): The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. *Information Systems Research*, 19(4), 417–433.

Reuters (2025): DeepSeek gives Europe's tech firms a chance to catch up in global AI race. https://www.reuters.com/technology/artificial-intelligence/deepseek-gives-europes-tech-firms-chance-catch-up-global-ai-race-2025-02-03/?utm_source=chatgpt.com (letöltve: 2025.02.06.).

Reuters (2025): Ride-hailing platform Lyft ties up with Anthropic for AI-powered customer care. https://www.reuters.com/technology/artificial-intelligence/ride-hailing-platform-lyft-ties-up-with-anthropic-ai-powered-customer-care-2025-02-06 (letöltve: 2025.02.07.).

Ringly (2025): 10 Real-Life Examples of Artificial Intelligence in 2025. https://www.ringly.io/blog/10-examples-of-artificial-intelligence-in-2025 (letöltve: 2025.02.07.).

Scaringi, J. A. – Mctaggart, R. A. – Alvin, M. D. – Atalay, M. – Bernstein, M. H. – Jayaraman, M. V. – Jindal, G. – Movson, J. S. – Swenson, D. W. – Baird, G. L. (2024): Implementing an AI algorithm in the clinical setting: a case study for the accuracy paradox. *European Radiology*. https://doi.org/10.1007/s00330-024-11332-z.

Shepardson, D. – Jin, H. (2021): U.S. opens probe into Tesla's Autopilot over emergency vehicle crashes, https://www.reuters.com/business/autos-transportation/us-opens-formal-safety-probe-into-tesla-autopilot-crashes-2021-08-16/ (letöltve: 2025.02.09.).

Szabó, H. (2023): A mesterséges intelligencia biztonsági kockázatai egy új korszak kezdetén. *Nemzetbiztonsági Szemle*, 11. évfolyam (2023) 4. szám 35–46., https://doi.org/10.32561/nsz.2023.4.3.

University of Cambridge (2022): Mathematical Paradox Demonstrates the Limits of AI. *University of Cambridge*. https://www.cam.ac.uk/research/news/mathematical-paradox-demonstrates-the-limits-of-ai (letöltve: 2025.02.09.).

Walz, E. (2024): NHTSA opens safety probe for up to 2.4M Tesla vehicles, https://www.automotivedive.com/news/nhtsa-opens-investigation-tesla-fsd-odi-crashes-autopilot/730353/ (letöltve: 2025.02.09.).

West, R. – Aydin, R. (2024): The Generative AI Paradox: ‚What It Can Create, It May Not Understand'. *arXiv preprint arXiv*:2311.00059. https://arxiv.org/abs/2311.00059 (letöltve: 2025. február 9.).

West, R. – Aydin, R. (2024): There and Back Again: The AI Alignment Paradox. *arXiv preprint arXiv:*2405.20806. https://arxiv.org/abs/2405.20806 (letöltve: 2025.02.09.).

Yin, Z. – Wang, H. – Horio, K. – Kawahara, D. – Sekine, S. (2024): Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. *arXiv:*2402.14531. https://arxiv.org/abs/2402.14531.

E cikk elkészítése során a szerző a ChaptGPT 4o és a Perplexity.AI Pro változatát használta ötletgeneráláshoz és forráskutatáshoz. A szerző az MI-alapú eszközök által generált szöveget közvetlenül nem használta, valamint valamennyi tényt és hivatkozást külön ellenőrzött.