

Jogpolitika

KAPPEL GERGELY*

Mesterségesintelligencia-alapú döntéstámogató és döntéshozó rendszerek kockázatai a vállalatok vezetői szintű döntéshozatalában: Szakirodalmi áttekintés

*Risks of Artificial Intelligence-Based Decision Support
and Decision-Making Systems in Executive-Level Decision-Making
in Companies: A Literature Review*

ABSZTRAKT

A tanulmány a mesterséges intelligencia (MI) alapú döntéshozó és döntéstámogató rendszerek vállalatok, valamint kis- és középvállalkozások vezetői szintű döntéshozatalában jelentkező kockázatait vizsgálja. A globális trendek és a digitális fejlődés hatására a vállalatok vezetésének egyre komplexebb döntéseket kell meghozniuk, amelyek támogatására kétségtelenül alkalmazhatók lehetnek az MI-alapú döntéshozó és döntéstámogató rendszerek. Mindazonáltal ez számos kockázatot rejtethet magában, így a tanulmány célja, hogy azonosítsa azokat a jogi, etikai és üzleti kockázatokat, amelyek az ilyen típusú MI rendszerek alkalmazásával kapcsolatban felmerülnek, különös tekintettel a vállalatvezetők döntéseire. Az elemzés egy szakirodalmi áttekintésre épül, amelyet végül a tanulmány összehasonlított a Stanford University által évente közzétett AI Index Reportokban található kérdőíves válaszokkal.

Tárgyszavak: mesterséges intelligencia, adatvédelem, automatizált döntéshozatal, kockázatkezelés

* Kappel Gergely, PhD-hallgató, Pécsi Tudományegyetem Állam- és Jogtudományi Kar Doktori Iskola, e-mail: kappel.gergely@pte.hu. A tanulmány a Kulturális és Innovációs Minisztérium ÚNKP-23-3-i. kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

ABSTRACT

The study examines the risks associated with artificial intelligence (AI) based decision-making and decision-support systems in the decision-making processes of company executives, as well as small and medium-sized enterprises. Due to global trends and digital advancements, company management increasingly faces complex decisions, which AI-based decision-making and decision-support systems may well be suited to support. However, this carries several risks, and the study aims to identify the legal, ethical, and business risks associated with the use of such AI systems, with a particular focus on the decisions made by company executives. The analysis is based on a literature review, which will ultimately be compared with survey responses found in the AI Index Reports published annually by Stanford University.

Keywords: Artificial Intelligence, data protection, automated decision-making, risk management

A világgazdaság összefonódásának és a globális trendeknek köszönhetően a vállalatok vezetésének egyre komplexebb, a fenntartható fejlődésnek és a pénzügyi stabilitásnak megfelelő döntéseket kell hoznia. A vállalati működés szempontjából a döntések során ez a komplex környezet változatos problémákat és kihívásokat támaszt, amit kétségtelenül tovább nehezítenek a tulajdonosi és más érdekelti körök által támasztott – cseppet sem kíméletes – elvárások is. Az elvárások sok esetben magas szintű nyomással párosulnak, amelyek következtében a rossz döntéshozatal a vállalat és végső soron a tulajdonosok pénzébe kerül. A menedzsmentszintű szegényes, nem megfelelő döntéshozatal átlagosan a nyereség legalább 3%-ába kerül a vállalatoknak. A digitális fejlődés által megteremtett lehetőségek, amilyen mondjuk a mesterséges intelligencia (MI) alapú döntéshozó és döntéstámogató rendszerek, nagyban hozzájárulhatnak a gyors és komplex döntéshozatalt igénylő akadályok minél zökkenőmentesebb abszolválásához. Nem csoda, hogy egyre több vállalat próbálja a mindennapi működésébe implementálni a különböző MI-alapú rendszereket.¹

Az IBM Institute for Business Value az Oxford Economicsszal közös kutatásában vizsgálták a vezérigazgatók döntéshozatalát a jelenkorunk legnagyobb ovációját körülölelő MI-vel kapcsolatban. A kutatás keretében megkérdeztek több mint 3000 vezérigazgatót (CEO), több mint 30 országot és 24 iparágat érintően. A kutatásból többek között az derül ki, hogy a CEO-k 43%-a használ valamilyen generatív MI-t a stratégiai döntéshozatalok során, illetve a vállalatok igazgatósági tagjainak 66%-a, míg a befektetői és hitelezői körök 64%-a ösztönzi a CEO-kat, hogy gyorsítsák fel az MI vállalati implementációját.² A Stanford University által közzétett 2024-es AI Index Reportból kiderül, hogy a befektetői érdeklődés évek óta meghatározó az MI-vel kapcsolatban, amely 2023-ban 189,2 milliárd USD volt globális vállalati befektetésként. Ebből 95,99 milliárd USD magánbefektetésként jelent meg a globális piacon.

¹ Purdy, Mark–Williams, A. Mark: How AI Can Help Leaders Make Better Decisions Under Pressure. <https://hbr.org/2023/10/how-ai-can-help-leaders-make-better-decisions-under-pressure> (2024. 05. 31.)

² IBM Institute for Business Value: CEO Decision-Making in the Age of AI. <https://www.ibm.com/downloads/cas/1V2XKXYJ> (2024. 05. 31.).

A befektetési piacot messzemenően az Egyesült Államok uralja a maga 67,22 milliárd USD-jével, míg második helyen Kína található 7,76 milliárd, harmadik helyen az Egyesült Királyság 3,78 milliárd, negyedik helyen pedig Németország 1,91 milliárd USD-vel. A befektetési számok – jóllehet a 2021-es³ befektetési csúcs óta visszaesés mutatkozik – alapján prognosztizálható, hogy az MI-alapú szoftverek és megoldások exponenciális növekedése tovább fokozódik majd a következő években. Az *AI Index Report*ból az is kiderül, hogy a generatív MI-be történő befektetések összege is kiugróan magas lett az elmúlt évben, hiszen míg 2022-ben 3 milliárd USD körül volt, addig 2023-ban már 25,23 milliárd USD volt globálisan a magánbefektetések száma a generatív MI területén. Ez tulajdonképpen a teljes MI-be történő magánbefektetések 26%-a, amely egy cseppet sem elhanyagolható arány.⁴

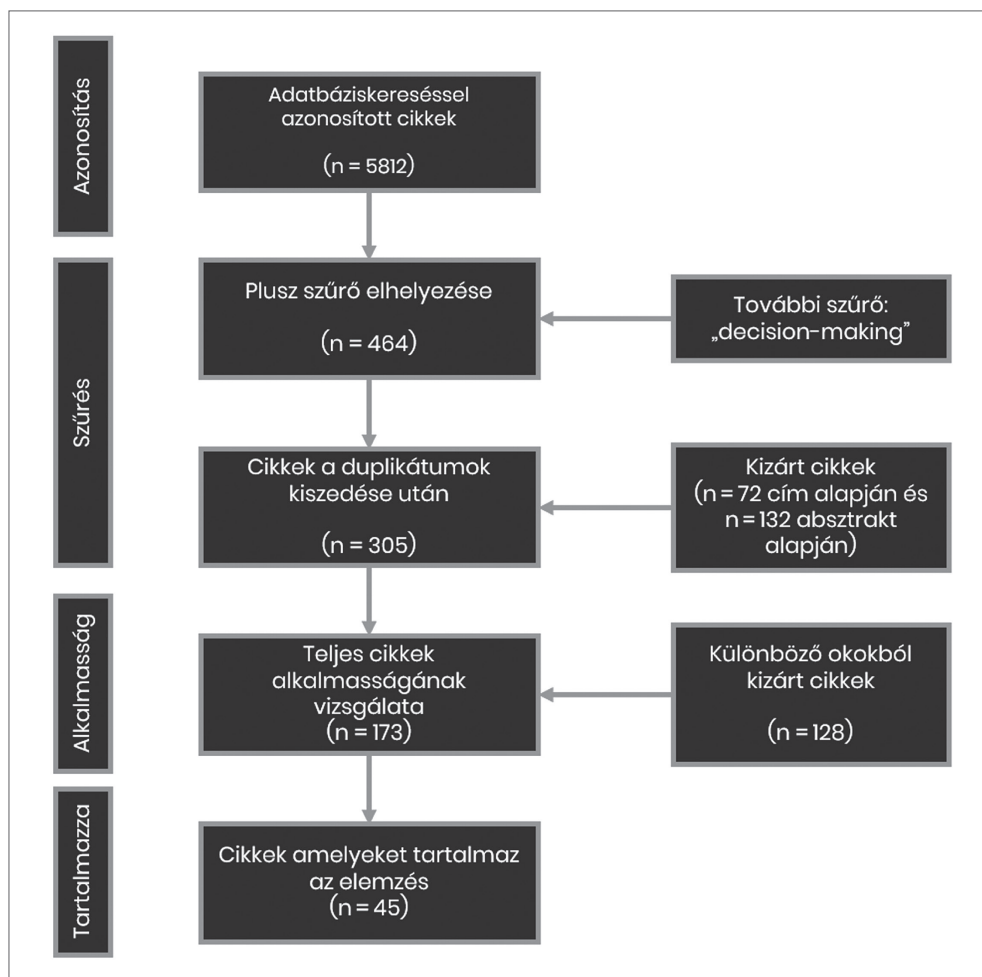
A generatív MI területén tapasztalható hatalmas befektetési növekedéssel párhuzamosan – nem meglepő módon – a vállalati felhasználás is növekszik. Természetesen nem csak a generatív MI megoldások száma bővül a vállalati alkalmazások terén, azonban kétségtelen, hogy ez tölti be a meghatározó szerepet – általánosságban – a következő üzleti években. A felhasználás növekedésével párhuzamosan emelkednek az MI-alapú megoldások kockázati esetei is. A kockázati esetkörök egy túlnyomó része meglátásom szerint az MI-alapú döntéshozó, illetve döntéstámogató rendszerek alkalmazása terén keletkezik. Mindazonáltal szükségesnek vélem a kockázati esetkörök (köztük is a leggyakrabban előforduló esetkörök) pontos meghatározását, amely hozzájárulhat a vállalatok minél gördülékenyebb felkészüléséhez a kockázatok mérséklése terén.

A kockázatok azonosításának több módja lehetséges, azonban kettőt mindenképp kiemelnék. Egyrészt lehetséges a kockázatok azonosítása azzal, hogy megvizsgáljuk az ezzel foglalkozó szakemberek, tudományos gondolkodók által készített szakirodalmakat, és ebből következtetünk a leggyakrabban vizsgált kockázatokra. Másrészt lehetséges a vállalatok, valamint kis- és középvállalkozások által már tapasztalt kockázatok kérdőíves azonosításával. E tanulmány mint szakirodalmi áttekintés az első azonosítási módot követi, azonban a szakirodalomban gyakran fellelhető kockázatokat egyúttal összehasonlítja az évente megjelenő *AI Index Report*-ban közölt és kérdőív alapján azonosított kockázatokkal. A tanulmány első része a kutatás menetét ismerteti, azon belül is az (i) adatforrásokat és kutatási stratégiát, valamint a (ii) kizárási kritériumokat. A második rész a kockázatok azonosítását ismerteti, így a (i) kockázatok azonosítását az átnézett szakirodalmak alapján; (ii) kockázatok vizsgálati prioritásának alakulását a vizsgált időszakban; és az (iii) eredmények összehasonlítását a vállalatok által megadott válaszok alapján. A harmadik rész olyan megoldási javaslatokat tartalmaz az ismertett kockázatokra, amelyek az áttekintett szakirodalmi művekben megjelennek. A negyedik rész a kutatással kapcsolatos következtetéseket tartalmazza és végül az ötödik fejezet a kutatás limitációit ismerteti.

³ 2021-ben a globális magánbefektetési szám közel 130 milliárd USD volt az elemzések szerint. Lásd MASLEJ, Nestor–FATTORINI, Loredana–PERRAULT, Raymond–PARLI, Vanessa–REUEL, Anka–BRYNJOLFFSSON, Erik–ETCHEMENDY, John–LIGETT, Katrina–LYONS, Terah–MANYIKA, James–NIEBLES, Juan Carlos–SHOHAM, Yoav–WALD, Russell–CLARK, Jack: *The AI Index 2024 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. 166. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf (2024. 05. 31.).

⁴ MASLEJ et al.: i. m., 242–244.

1. A kutatás célja és menete



1. ábra. A kutatás menetének lépései

A kutatás célja, hogy azonosítsam jogi, etikai és üzleti szempontból azokat a kockázatokat, amelyek az MI-alapú döntéshozó és döntéstámogató rendszerek használatából származnak. Az előbbieken belül is a vállalatok, és némiképp a kis- és középvállalkozások vezetői szintű döntéshozatalára fókuszálva. Tekintettel arra, hogy a vizsgált szakirodalmi művek eltérő jogrendszereket, valamint sajátos definíciókat alkalmaznak, így a jelen tanulmányban használt vállalat, valamint kis- és középvállalkozás meghatározását mellőzöm. Így az nem követi a magyar hatályos jogban meghatározott definíciókat, sőt az európai uniós joganyagokban fellelhető kritériumokat, minősítési szempontokat sem. A kutatás és a jelen tanulmány megírásakor ezért absztrakt megközelítésben használom a vállalatokat, valamint a kis- és középvállalkozásokat. Ezzel együtt a vizsgálatomat szintén nem a magyar hatályos jog-

szabályok lencsében keresztül, hanem egy, a kutatás célját – álláspontom szerint – leginkább támogató absztrakciós szinten végzem. Mindazonáltal a kockázatok megfelelő azonosításával és ezek kezelésével megelőzhetők a vállalkozások reputációs és anyagi kárai, továbbá a felelősségi szerepek is jobban elhatárolhatók a különböző jogviszonyok tekintetében. Voltaképpen a kutatás céljaként feltehető kérdés, hogy melyek azok a leggyakrabban előforduló kockázatok, amelyeket a szakirodalom azonosított az MI-alapú döntéshozó és döntéstámogató rendszerek tekintetében. A kutatás során arra törekedtem, hogy feltárjam a mások által, szakirodalmi szinten azonosított kockázatok és megoldási javaslatokat, amelyeket végül ütköztetek a megkérdezett szervezetek által vélt kockázatokkal. Ezáltal a jelen tanulmányban véleményeket, és nem sajátos javaslatokat fogalmazok meg.

A kutatás menete 5 részből állt (lásd 1. ábra), amelynek első részében címszavakat alakítottam ki úgy, hogy az egyes címszavak szinonimájához az *OpenAI* által biztosított *ChatGPT* szolgáltatást használtam. A második részében a kialakított címszavak alapján keresést folytattam le a *Scopus* és a *Web of Science* adatbázisban, tekintettel arra, hogy ez a két adatbázis volt alkalmas arra, hogy a későbbiekben ismertetett szűrési kritériumokat, valamint az összetett kulcsszavas keresést megfelelően használjam. Kezdetben a fentiekben megjelölt témát tágabb viszonyrendszerben szándékoztam vizsgálni, azonban a találatok nagy mennyisége miatt további szűrőt is alkalmaztam, így a vizsgálat területét is szűkítettem. A kutatás harmadik részében kiszűrtem a duplikátumokat. Ezután az absztraktok vizsgálata következett, amely alapján tovább szűrtem a kutatás szempontjából nem releváns cikkeket az 1.2. cím alatt részletezett kizárási kritériumok alapján. Ötödik lépésként áttekintettem a megmaradt cikkeket, amelyeket egyúttal tovább szűkítettem az *open access* szakirodalmi művekre.

1.1. Adatforrások és kutatási stratégia

A jelen tanulmány alapját képező kutatást az alábbi szempontok és stratégia mentén alakítottam ki, amelynek első lépése volt egy átfogó címszavakból álló lista⁵

⁵ A címszavakból álló lista első vaglyagos elemei: *Legal; regulation; Law; Contract law; Consumer law; Business law; Corporate law; Employment law; Intellectual property law; Tax law; Environmental law; Regulatory compliance; Privacy law; Cybersecurity law; Commercial litigation; Antitrust law; Banking and finance law; Real estate law; Labor law; Mergers and acquisitions; Bankruptcy law; Intellectual property rights; Data protection; Employment contracts; Trademark law; Copyright law; Product liability; Healthcare law; Estate planning; Antitrust regulations; Immigration law; Legal compliance; Taxation; Insurance law; Securities regulation; Environmental compliance; Corporate governance; Contract negotiations; Regulatory affairs; Trade secrets; Compliance management; Employment disputes; Data privacy; Licensing agreements; Labor disputes; Financial regulations; Corporate transactions; Business contracts; Intellectual property protection; Tax compliance; Legal risk management; Contract disputes; Legal counsel; Corporate litigation; Corporate governance; Legal compliance; Corporate ethics; Corporate responsibility; Legal advisory; Corporate policies; Legal department; Legal regulations; Corporate legal framework; Legal issues in business; Legal risk assessment.* A címszavakból álló lista második vaglyagos elemei: *company; corporation; firm; business; enterprise; organization; concern; institution; agency; establishment; venture; house; conglomerate; consortium; partnership; agency; firm.* A címszavakból álló lista harmadik vaglyagos elemei: *Artificial Intelligence; AI.*

azonosítása. A kutatás során interdiszciplináris megközelítést alkalmaztam, így a keresés alatt nem korlátoztam a szakirodalmi művek szelekcióját jogtudományi alapon. Az alkalmazott két adatbázisban, azaz a *Scopus*-ban és a *Web of Science*-ben a keresést 2023. november 3. napján végeztem el. A *Scopus* esetében a következő szűrési kritériumokat alkalmaztam: (i) végleges; (ii) angol nyelvű; (iii) publikációk, könyvek, könyvrészletek és ismertető/kritikák. A *Web of Science* esetében pedig a következőket: (i) angol nyelvű; (ii) publikációk, könyvek és ismertető/kritikák. A vizsgálat időbeli terjedelme mindkét adatbázisban a 2013-tól 2023-ig tartó megjelenéseket foglalta magába. Az így kapott szakirodalmi lista 5812 találatból állt, amely a kutatás ezen szakaszában még egy tágabb azonosítási célt szolgált. Felismerve, hogy a tágabb kutatási téma szakirodalmi száma az utóbbi években exponenciálisan nőtt, és ez a tendencia a mai napig tart, így a vizsgálathoz egy további címszót, nevezetesen a „*decision-making*”-et alkalmaztam szűrésként. Ennek köszönhetően 464 találatra szűkítettem a listát.

A szűrési folyamat következő nagyobb fázisa a talált szakirodalom cím alapján történő további szűkítése volt (az ezzel kapcsolatos kizárási kritériumokat lásd az 1.2. cím alatt). A cím alapján történő szűrés után a találat 392 szakirodalmi műre korlátozódott, amely a duplikátumok kiszűrése után 305 találatra korlátozódott. Ezt követte az absztraktok átolvasása alapján történő további szűrés (kizárási kritériumokat lásd 1.2. cím alatt), amelynek eredményeként 173 szakirodalmi műre korlátozódott az irodalmi áttekintés. A 173 találat teljes átnézése során további olyan szakirodalmi művek estek ki a vizsgálat alól, amelyek nem voltak nyílt forrásúak vagy a hozzáférés korlátozott volt, és így nem volt hozzá jogosultságom. Végül 137 szakirodalmi művet vizsgáltam át részletesen, amelyből végezetül 45 szakirodalmi mű bizonyult a kutatási céloom szempontjából relevánsnak.

A kutatás utolsó fázisaként pedig a 45 szakirodalmi műből kiszűrtem, hogy milyen kockázatokat vizsgáltak a leggyakrabban a szerzők, majd ebből levontam a következtetéseimet, amely a jelen tanulmány 2. fejezetében került kifejtésre.

1.2. Kizárási kritériumok

A kutatás során kizárási kritérium volt minden olyan szakirodalmi mű, amely a közszférában, igazságszolgáltatásban alkalmazott MI-alapú rendszerekkel foglalkozott. Továbbá kizárás tárgya volt minden olyan mű, amely tisztán informatikai, matematikai, fizikai stb. megfontolásokat, vizsgálatokat tartalmazott. Végül kizártam az önvezető autókat, önvezető rendszereket vizsgáló szakirodalmi műveket. Kizárólag azokat a cikkeket, könyveket stb. vizsgáltam, amelyek az MI-alapú döntéshozó és döntéstámogató rendszerek kockázatait akképpen elemezték, hogy az vállalati, valamint kis- és középvállalkozói szektorban jelen lévő cégek vezetői döntéseinek támogatása céljából relevanciával bír. Így a vizsgált művek között megtalálható olyan is, amely az egészségügyi szférát érinti, ellenben egyúttal olyan kockázatokat azonosít, amelyek a vállalati, valamint a kis- és középvállalkozói szintű döntéshozatal során is meghatározóak tudnak lenni.

2. A kockázatok azonosítása

Ebben a fejezetben a jelen tanulmány azokat a kockázatokat ismerteti, amelyek az áttekintett szakirodalmi művek esetében a leginkább megjelentek az MI-alapú eszközökkel kapcsolatban. A vizsgált művek kutatási területei eltérő képet mutatnak, így az áttekintés céljából a tanulmány függeléke tartalmazza az egyes művek vizsgálati területét, valamint a megjelenés dátumát. Az alábbiakban ismertetem az egyes kockázatokat, illetve azok fogalmát a talált szakirodalmakban, illetve egyúttal a vizsgálatuk gyakoriságának előfordulását is a szakirodalmi művek megjelenési évei alapján. A kockázatokat az alapján azonosítottam, hogy az egyes művek valamilyen szinten vizsgálják az adott mű keretében, azonban ahol szimplán említik csak az adott kockázatot, azokat figyelmen kívül hagytam. A feltárt kockázatok között szerepelnek olyanok is, amelyek priméren nem jogi kockázatok, azonban másodlagosan azzá válhatnak, ilyen például a pontatlanság esetköre, amely kárt okozhat a vállalatok, valamint a kis- és középvállalkozások számára, így azok rendezése jogi eszközöket igényel.

2.1. A kockázatok azonosítása

Az áttekintett szakirodalmi művek alapján a következő kockázatok, illetve kockázati területek jelentek meg a leggyakrabban: (i) elfogultság; (ii) diszkrimináció; (iii) igazságtalanság; (iv) átláthatóság; (v) megmagyarázhatóság; (vi) értelmezhetőség; (vii) érthetőség; (viii) megbízhatóság; (ix) bizalomhiány; (x) adatvédelem; (xi) kibervédelem; (xii) hozzáférhetőség az adatokhoz; (xiii) pontatlanság; (xiv) stabilitás; (xv) elszámoltathatóság; (xvi) a jogi keretrendszer hiánya. Az egyes kockázatok sokszor átfedésbe kerülnek a szakirodalmi művekben, így előfordul, hogy egyes kockázatot más kockázat elemeként vagy annak szinonimájaként vagy éppenséggel egy másik kockázat következményeként azonosítanak.

2.1.1. Elfogultság

Az elfogultság definícióját a szakirodalmi művekből mellőzik, azt tulajdonképpen evidenciával kezelik. Azonban van olyan, amelyben akként hivatkoznak rá, mint hogy valaki/valami igazságtalanul előítéletes valaki/valami mellett vagy ellen.⁶ Az MI-alapú döntéshozó és döntéstámogató rendszerek esetében az elfogultságnak több megjelenési formáját ismerjük. Így értekezhetünk a tanítási adatok alapján kialakult elfogultságról,⁷ a bemeneti adatok alapján kialakult elfogultságról,⁸ a fejlesztő által az

⁶ CHILLAR, D.–AGUILERA, R. V.: An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research. *Business & Society*, 2022/5, 1211. (<https://doi.org/10.1177/00076503221080959>).

⁷ SCHÖNBERGER, Daniel: Artificial Intelligence in Healthcare: a Critical Analysis of the Legal and Ethical Implications. *International Journal of Law and Information Technology*, 2019/2, 176. (<https://doi.org/10.1093/ijlit/eaz004>). JOHNSON, Kristin N.: Automating the Risk of Bias. *George Washington Law Review*, 2019/6., *Tulane Public Law Research Paper* No. 19–12, 2019, 1222.

⁸ JOHNSON: i. m., 1240.

algoritmusba táplált elfogult változó következtében keletkező elfogultságról,⁹ történelmi és szisztematikus elfogultságról,¹⁰ kognitív elfogultságról¹¹ stb. Az elfogultság mint kockázat jelenik meg legtöbbször a vizsgált szakirodalmi művekben, nevezetesen 45 műből 36 foglalkozik vagy érinti ezt a problémát (lásd: 2. ábra).

2.1.2. Diszkrimináció

Ahogy a fentiekben, úgy a diszkrimináció aspektusából is különféle kockázatok és megközelítési módok jelennek meg a szakirodalmi művekben. Egyrészt az algoritmus kialakítása során érhető tetten a diszkriminatív tervezési hibákból származó hátrányos megkülönböztetés mint kockázati tényező.¹² Más szerzők azt hangsúlyozzák, hogy MI-alapú rendszereknél a diszkrimináció melegágya a diszkriminatív bemeneti adatok alkalmazása.¹³ Mások pedig arról értekeznek, hogy a gépi tanulással kapcsolatos szükségszerű szubjektív döntések következménye az MI-alapú döntéshozó és döntéstámogató rendszerek diszkriminatívvá válása. E körben említik a képzési adatok gyűjtésének, kezelésének módját, a modell kialakítását stb.¹⁴

2.1.3. Igazságtalanság

Az igazságtalanság kérdésköre túlnyomórészt tág értelemben jelenik meg az áttekintett szakirodalmi művekben. Egyes helyeken az igazságosság megvalósulását azért látják kiemelten fontosnak, hogy az MI-alapú rendszer elfogultságát és diszkriminatív hatását meg lehessen előzni.¹⁵ Egyes szerzők azokat az ember által fejlesztett elemeket vizsgálják, amelyek befolyásolják az MI-rendszer igazságosságát. Ilyen a módszer, amit az MI rendszer létrehozásához használnak (például a tanulási modell stb.), ilyen az algoritmus, valamint a fizikai technológiai háttér.¹⁶ Az előbbieken túl számos más aspektusból is vizsgálják az igazságosság kérdését az áttekintett szakirodalmi művek.

⁹ KRUPPIY, Tetyana (Tanya): A Vulnerability Analysis: Theorising the Impact of Artificial Intelligence Decision-Making Processes on Individuals, Society and Human Diversity from a Social Justice Perspective. *Computer Law & Security Review*, 2020, 38. sz. 6. (<https://doi.org/10.1016/j.clsr.2020.105429>).

¹⁰ MARDA, Vidushi: Artificial Intelligence Policy in India: a Framework for Engaging the Limits of Data-Driven Decision-Making. *Philosophical Transactions of the Royal Society A* 376: 20180087, 2018, 7. (<http://dx.doi.org/10.1098/rsta.2018.0087>).

¹¹ JOHNSON: i. m., 1252.

¹² KIESOW CORTEZ, E.–MASLEJ, N.: Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA. *European Journal of Risk Regulation*. 2023/3, 463. (DOI: 10.1017/err.2023.61).

¹³ BEKKUM, Marvin van–BORGESIU, Frederik Zuiderveen: Using Sensitive Data to Prevent Discrimination by Artificial Intelligence: Does the GDPR Need a New Exception? *Computer Law & Security Review*, 2023, 3. (<https://doi.org/10.1016/j.clsr.2022.105770>).

¹⁴ SELBST, Andrew D.: Negligence and AI's Human Users. *Boston University Law Review*, 2020/4, 1354.

¹⁵ TAN, Lorwai–TIVEY, David–KOPUNIC, Helena–BABIDGE, Wendy–LANGLEY, Sally–MADDERN, Guy: Part 1: Artificial Intelligence Technology in Surgery. *ANZ Journal of Surgery*, 2020/12, 2411. (<https://doi.org/10.1111/ans.16343>).

¹⁶ TSCHIDER, Charlotte: Beyond the Black Box. *Denver Law Review*, 2021/3, 692.

2.1.4. Átláthatóság

Az átláthatóság kérdése a második leginkább vizsgált terület az MI-alapú döntéshozó és döntéstámogató rendszerek kockázatait tekintetében az áttekintett művek alapján. Az átláthatóság fogalmát egységesen nehéz meghatározni, így e tekintetben is eltérnek az egyes művek megközelítései. Tovább bonyolítja a helyzetet, hogy az átláthatóság a címzettek esetében is eltérő fogalommal bír. Példának okáért egyes szerzők szerint a fejlesztő szemszögéből az átláthatóság egyfajta megértése annak, hogy az algoritmus megfelelően működik-e, annak érdekében, hogy a felmerülő hibákat, ellentmondásokat képes legyen feloldani. A felhasználó oldaláról pedig az a körülmény, tulajdonság, hogy a rendszer mit és miért csinál, mi vezetett az adott döntés meghozatalához, stb.¹⁷ Mások az MI-alapú rendszer felé táplált bizalom alap-elemeként tekintenek az átláthatóság megvalósulására, valamint az érintettek tájékoztatásának oldaláról közelítik meg azt.¹⁸

2.1.5. Megmagyarázhatóság

A megmagyarázhatóság körében kiemelt vizsgálati szempont az úgynevezett *black-box* hatás, valamint az, hogy a megfelelően szervezett MI-rendszer megmagyarázhatósága kiemelten növeli a felhasználói szintű bizalmat. Továbbá szorgalmazza a döntéshozatal azonosításának folyamatát, így képesek vagyunk értelmezni az adott kimenet okait.¹⁹

2.1.6. Értelmezhetőség

Az értelmezhetőség mint kockázat szintén szorosan kapcsolódik az MI-alapú rendszerrel kapcsolatos *black-box* hatáshoz. A kockázat abban keresendő, hogy sokszor az MI-modell működésének megértése problémákba ütközik. Voltaképpen a döntés technikai megvalósulásához kapcsolódik. Ahogy a megmagyarázhatóság, így ez is alapeleme a megbízható MI kialakításának.²⁰ Továbbá az értelmezhetőség növelésével egyúttal növeljük az MI-rendszer átláthatóságát is, így ennyiben ez a két kockázat is összekapcsolódik egymással.²¹

¹⁷ FELZMANN, Heike–VILLARONGA, Eduard Fosch–LUTZ, Christoph–TAMÓ–LARRIEUX, Aurelia: Transparency You Can Trust: Transparency Requirements for Artificial Intelligence Between Legal Norms and Contextual Concerns. *Big Data & Society*, 2019/1., 5. (<https://doi.org/10.1177/2053951719860542>).

¹⁸ DARA, Rozita–HAZRATI FARD, Seyed Mehdi–KAUR, Jasmin: Recommendations for Ethical and Responsible Use of Artificial Intelligence in Digital Agriculture. *Frontiers in Artificial Intelligence*, 2022, 4. (DOI: 10.3389/frai.2022.884192).

¹⁹ ALI, Sajid–ABUHMED, Tamer–EL-SAPPAGH, Shaker–MUHAMMAD, Khan–ALONSO-MORAL, Jose M.–CONFALIONERI, Roberto–GUIDOTTI, Riccardo–SER, Javier Del–DÍAZ-RODRÍGUEZ, Natalia–HERRERA, Francisco: Explainable Artificial Intelligence (XAI): What We Know and What is Left to Attain Trustworthy Artificial Intelligence. *Information Fusion*, 2023, 2. és 6. (<https://doi.org/10.1016/j.inffus.2023.101805>).

²⁰ ALI et al.: i. m., 8.

²¹ RIZINSKI, Maryan–PESHOV, Hristijan–MISHEV, Kostadin–CHITKUSHEV, Lubomir T.–VODENSKA, Irena–TRAJANOV, Dimitar: Ethically Responsible Machine Learning in Fintech. *IEEE Access*, 2016, 11.

2.1.7. Érthetőség

Az érthetőség hiánya mint kockázat szintén a *black-box* hatásból származik, tekintettel arra, hogy sokszor a rendszerek tervezői se tudják ismertetni a döntések mögött meghúzódó pontos indokokat.²² Hasonló kockázati kategória, mint a megmagyarázhatóság vagy az értelmezhetőség hiányából származó kockázati kategória, azonban az áttekintett szakirodalom mégis külön értékeli őket.

2.1.8. Megbízhatóság

Egyes szerzők szerint a megbízhatóság arra vonatkozik, hogy az MI-alapú rendszer jelezze, mikor fog meghibásodni, működésképtelenné válni.²³ Mások abban a kontextusban érintik, hogy az algoritmus döntéseinek megbízhatónak kell lenniük, különösen akkor, amikor kritikus döntéseket kell meghozni, például tőzsdei vásárlások végrehajtása.²⁴ Amennyiben a kimeneti megbízhatóság oldaláról vizsgáljuk, akkor a megbízhatóság erőteljesen ki van téve a felhasznált adatok mennyiségének és minőségének, így ezek hiánya oda vezethet, hogy az MI-alapú döntéshozó és döntéstámogató rendszer kimeneti előrejelzése nem lesz megbízható.²⁵

2.1.9. Bizalomhiány

Az MI-alapú rendszerek felé táplált bizalom kialakítása kurrens téma a szakirodalom számára, mint amit szükségszerűen el kell érni ahhoz, hogy ezek a rendszerek betöltsék szerepüket. Több szerző így egy elérendő célként határozza meg a bizalmat, amelynek eleme például a jogszerűség, etikusság, valamint technikai és társadalmi megbízhatóság.²⁶ Vannak szerzők, akik pedig azt vizsgálják, hogy az átláthatóság, azaz hogy a felhasználók milyen fokú képet kapnak az adott MI-alapú rendszerről, nagyban meghatározza a rendszer felé táplált bizalmat.²⁷ Értelmszerűen felhasználói oldalról, például egy vállalat által a vezetői szintű döntéshozatalba történő implementálásakor kiemelt szerepet kap, hogy a vállalat vezetői milyen mértékben bíznak az adott rendszerben, azt mennyire sikerül a mindennapi döntéshozatal részévé tenni. Amennyiben az MI-rendszer kialakítása nem szorgalmazza a bizalom megteremtését, úgy a vállalat számára költség és innováció oldalról ez negatív következményként jelentkezhet. Előfordulhat tehát, hogy egy vállalati döntés alapján integrálják a menedzsment részére az adott MI-rendszert, azonban az azzal kapcsolatos bizalom hiánya oda vezethet, hogy a rendszer tényleges

²² SCHÖNBERGER: i. m., 177.

²³ DARA et al.: i. m., 5.

²⁴ BUI, Toan Huu–NGUYEN, Van Phuoc: The Impact of Artificial Intelligence and Digital Economy on Vietnam's Legal System. *International Journal for the Semiotics of Law*, 2023, 981. (<https://doi.org/10.1007/s11196-022-09927-0>).

²⁵ TSCHIDER: i. m., 693.

²⁶ EROĞLU, Muzzafer–KARATEPE KAYA, Meltem: Impact of Artificial Intelligence on Corporate Board Diversity Policies and Regulations. *European Business Organization Law Review*, 2022, 547. (<https://doi.org/10.1007/s40804-022-00251-5>).

²⁷ ALI et al.: i. m., 32.

alkalmazása elmarad. Így az MI-rendszer implementálásakor meghatározó kockázat az, hogy egy bizalmat szorgalmazó vagy azt hátráltató MI-rendszer kerül beépítésre a vezetői döntéshozatalba.

2.1.10. Adatvédelem

E körben a szakirodalmi művek legtöbbször a GDPR által meghatározott keretrendszert vizsgálják,²⁸ illetve ezzel kapcsolatban a személyes adatok érvényesülését az MI-alapú döntéshozó és döntéstámogató rendszerek tekintetében. Mindazonáltal megjelennek más adatvédelmi szabályozásból származó keretrendszerek is e témakörben. Továbbá ebben a kockázati témakörben érvényesül a nem megfelelő adatmenedzsment kialakítása is mint kockázat.²⁹ Az adatvédelemből származó kockázatok egyes művek szerint a nagy mennyiségű felhasznált adatokból keletkeznek, hiszen az MI-alapú döntéshozó és döntéstámogató rendszerek többségében nagy adatbázisok alapján predesztinálnak.³⁰

2.1.11. Kibervédelem

A kibervédelem mint kockázat is magától értetődő ezeknél az MI-alapú rendszereknél, tekintettel arra, hogy nagy mennyiségű adatot használnak fel a működéshez, és értelemszerűen – igaz szektorálisan eltérő módon, de – ezek között az adatok között sokszor található érzékeny, adott esetben különleges adat vagy más üzleti titkot tartalmazó információ. Példának okáért egyes szerzők az egészségügyi szektor tekintetében vizsgálják és emelik ki a kibervédelem fontosságát. Hangsúlyozzák, hogy a kockázat relevanciája kiemelkedően magas, hiszen például az Egyesült Államokban körülbelül 45 millió embert érintett egészségügyi kibertámadás 2021-ben.³¹

2.1.12. Hozzáférhetőség az adatokhoz

Tekintettel arra, hogy az MI-alapú döntéshozó és döntéstámogató rendszerekhez óriási mennyiségű adatra van szükség, így kiemelt szerepet kap az adatokhoz való hozzáférés kérdése. A megfelelő típusú és mennyiségű adattal más kockázatok előfordulási aránya is csökkenthető, például az elfogultság vagy a pontosság mértéke. Sőt adott esetben az adatokhoz való hozzáférés hiánya könnyen diszkriminatív döntéshez is vezethet. Ilyen például India esete, ahol a magánszektorban kevésbé fér-

²⁸ BEKKUM: i. m.; LEE, Michelle Seng Ah–COBBE Jennifer–JANSSEN, Heleen–SINGH, Jatinder: Defining the Scope of AI ADM System Risk Assessment. In: Kosta, Eleni–Leenes, Ronald–Kamara, Irene (eds.): *Research Handbook on EU Data Protection Law. Research Handbooks in European Law*. Edward Elgar Publishing, Cheltenham, 2022, 405–434. (<https://doi.org/10.4337/9781800371682.00025>).

²⁹ HILB, Michael: Toward Artificial Governance? The Role of Artificial Intelligence in Shaping the Future of Corporate Governance. *Journal of Management and Governance*, 2020, 859. (<https://doi.org/10.1007/s10997-020-09519-9>).

³⁰ IQBAL, Javed–CORTÉS JAIMES, Diana Carolina–MAKINENI, Pallavi et al.: Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine. *Cureus*, 2023/9, 8. (DOI: 10.7759/cureus.44658).

³¹ VERITTI, Daniele–RUBINATO, Leopoldo–SARAO, Valentina et al.: Behind the Mask: a Critical Perspective on the Ethical, Moral, and Legal Implications of AI in Ophthalmology. *Graefes Arch Clin Exp Ophthalmol*, 2024/3, 2. (<https://doi.org/10.1007/s00417-023-06245-4>).

nek hozzá más piaci vagy közadatbázisokhoz. Ebből következően a hátrányos helyzetű csoportokhoz tartozó személyek akár nemi, kaszt vagy földrajzi elhelyezkedés alapján diszkriminatív döntés elszenvedői lehetnek.³²

2.1.13. Pontatlanság

A pontosság kérdésköre, illetve az annak hiányából származó kockázatok is kiemelt helyen vannak az áttekintett szakirodalmi művekben. A vizsgált művek több mint fele érinti ennek a kockázatnak a hatását, kapcsolódását más kockázatokkal. Például egyes szerzők arról értekeznek, hogy az elfogult képzési adatkészlet az MI-rendszer kimenetének pontosságát nagyban befolyásolja, továbbá a gépi tanuláson alapuló modellek helytelen képzése pontatlan eredményekhez vezethet.³³ Magától értetődően a pontatlan döntés negatív következményekkel járhat egy MI-alapú döntéshozó vagy döntéstámogató rendszernél, amely – ahogy más kockázatoknál is – egyúttal járhat vagyoni és reputációs kárral egyaránt egy vállalat vagy egy kis- és középvállalkozás számára.

2.1.14. Stabilitás

Az áttekintett szakirodalmi művek több helyen is foglalkoznak az MI-rendszer stabilitásával mint szükségeszerű alapelvvvel, azonban annak definíciós meghatározására vagy elemeinek kifejtésére többnyire nem térnek ki. Azonban egyes szerzők a stabilitással kapcsolatban kifejtik, hogy azalatt az adott rendszer azon tulajdonságát értik, hogy meg tudja őrizni a teljesítménye minőségét változó körülmények között is.³⁴ Tehát a stabilitás hiánya ezáltal kapcsolatban áll a rendszer megbízhatóságának hiányával is.

2.1.15. Elszámoltathatóság

Az elszámoltathatóság kérdéskörét egyes szakirodalmi művek az átláthatósággal együtt tárgyalják, mint utóbbi feltétele az előbbi érvényesülésének.³⁵ Az elszámoltathatóság elveként jelenik meg, hogy az MI-alapú döntéshozó és döntéstámogató rendszerek képesek vagy képesnek kellene lenniük arra, hogy a kimeneti döntést kvázi meg tudják magyarázni, illetve ismertetni tudják a mögötte meghúzó indokokat.³⁶ A felelősség, illetve a felelősségre vonás szempontjából meghatározó az elszámoltathatóság érvényesülése az ilyen típusú MI-alapú rendszereknél, hiszen a döntéseik hatással lehetnek közvetlenül az egyénekre is.³⁷

³² MARDÁ: i. m., 7.

³³ VERITTI et al.: i. m., 2.

³⁴ ALI et al.: i. m., 37.

³⁵ MARDÁ: i. m., 9.

³⁶ KRAUSOVÁ, Alžběta–MORAVEC, Václav: Disappearing Authorship: Ethical Protection of AI-Generated News, *JIPITEC*, 2022/2, 134.

³⁷ ROKSHAD, Rata–DUCRET, Maxime–CHAURASIA, Akhilanand–KARTEVA, Teodora–RADENKOVIC, Miroslav–ROGANOVIC, Jelena–HAMDAN, Manal–MOHAMMAD-RAHIMI, Hossein–KROIS, Joachim–LAHOUD, Pierre–SCHWENDICKE, Falk: Ethical Considerations on Artificial Intelligence in Dentistry: A Framework and Checklist. *Journal of Dentistry*, 2023, 3. (<https://doi.org/10.1016/j.jdent.2023.104593>).

2.1.16. Jogi keretrendszer hiánya

A jogszabályi keretrendszer hiánya vagy nem megfelelő volta magától értetődő módon kockázatként jelenik meg az MI-alapú döntéshozó és döntéstámogató rendszerek alkalmazása terén. Mindazonáltal az áttekintett szakirodalmi művekben viszonylag ritkán jelenik meg mint ekként értékelt terület. A 45 áttekintett műből összesen 4 műben foglalkoztak ezzel a kockázati tényezővel. A megfelelő jogi keretrendszerrel elősegíthető a bizalom növelése az adott MI-rendszer használatával kapcsolatban, érte ezalatt a használatából származó felelősségi kérdések konkretizálását is.³⁸

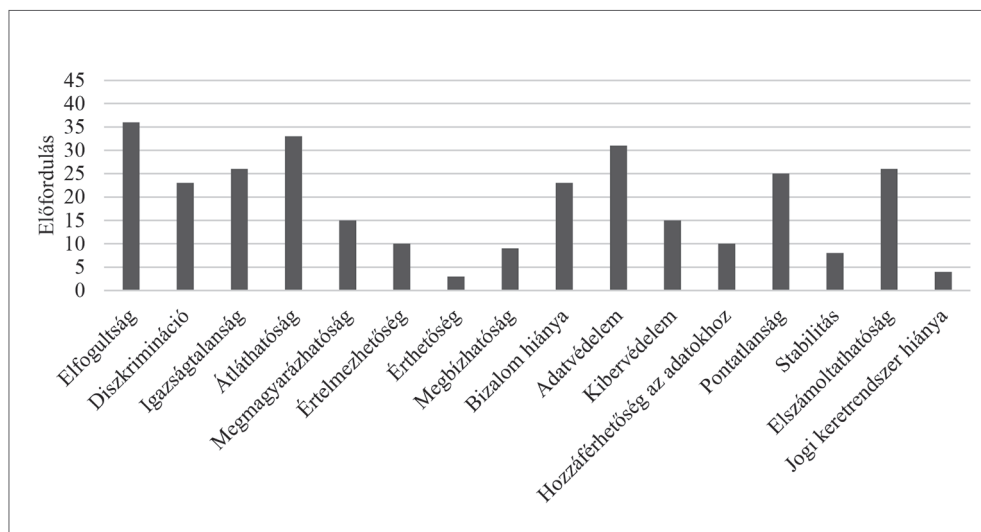
2.2. Kockázatok alakulása

A fenti kockázatok a vizsgált időszakban eltérő módon jelentek meg az áttekintett művekben. A három legtöbbet érintett kockázat az (i) elfogultság; (ii) átláthatóság; (iii) adatvédelem. Mindazonáltal szükséges megemlíteni, hogy egyes kockázatok a szerzők olykor más kockázatnak az alkategóriájaként értékelték, így például a megbízható (értsd ez alatt az MI-vel kapcsolatos bizalmat) MI-alapú rendszer eleme az igazságosság, az átláthatóság, az értelmezhetőség, a megmagyarázhatóság és a stabilitás.³⁹ Meglátásom szerint előfordul az is, hogy az egyes kockázati megnevezéseket szinonimaként, és nem eltérő kockázatként értékelnek. Ebből kiindulva a kockázatok vizsgálati előfordulásának arányát nehéz objektív szempontból vizsgálni. Mindazonáltal a 2. ábra tartalmazza a 45 szakirodalmi műben az egyes kockázatok vizsgálatának előfordulási számát. A vizsgált szakirodalomban a legkorábban publikált mű 2016-os, míg a legújabb 2023-as.

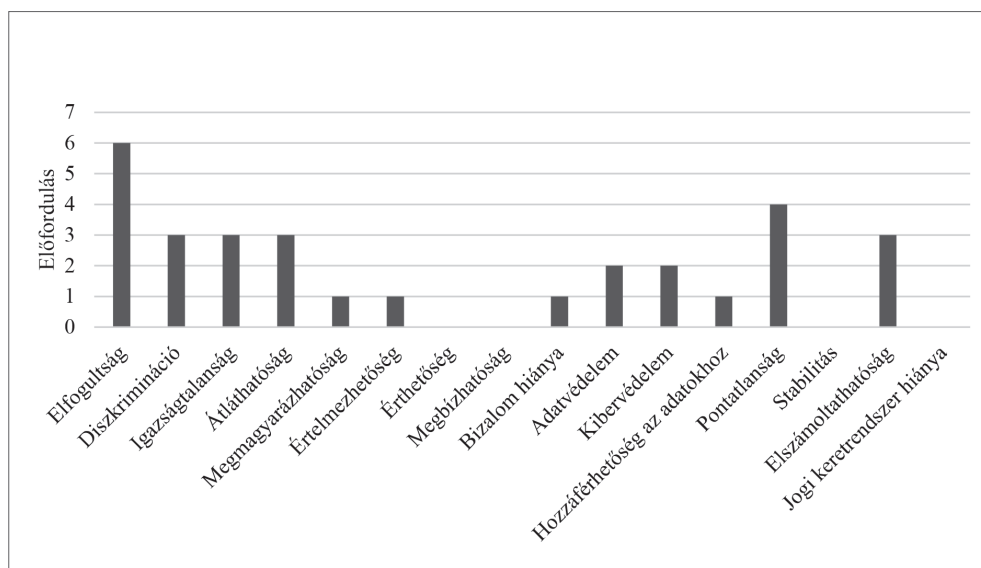
Tekintettel arra, hogy a legtöbb átnézett szakirodalmi mű 2020 és 2023 között jelent meg, így a kockázatok alakulását az alábbiakban ezekre a művekre tekintettel prezentálom. Láthatjuk a 3–6. ábrák alapján, hogy az elfogultság kérdésével foglalkoztak leginkább a szerzők. Ezt követte az átláthatóság kérdése, és olykor az adatvédelemből származó kockázatok, jogi vagy etikai szempontból. Meglepő módon a kibervédelemmel kapcsolatos kockázatok kevésbé jelentek meg, de ugyanígy kevés esetben foglalkoztak az adatokhoz való hozzáféréssel (értsük ez alatt a fennálló adatbázisokhoz vagy általánosságban a tanításhoz stb. szükséges adatokhoz való hozzáférést) vagy az értelmezhetőséggel kapcsolatos kockázatokat.

³⁸ HUANG, Zhilian–GEORGE, Mithun Mohan–TAN, Yi-Roe–NATARAJAN, Karthiga–DEVASAGAYAM, Emily–TAY, Evonne–MANESH, Abi–VARGHESE, George M.–ABRAHAM, Ooriapadickal Cherian–ZACHARIAH, Anand–YAP, Peiling–LALL, Dorothy–CHOW, Angela: Are Physicians Ready for Precision Antibiotic Prescribing? A Qualitative Analysis of the Acceptance of Artificial Intelligence-Enabled Clinical Decision Support Systems in India and Singapore. *Journal of Global Antimicrobial Resistance*, 2023, 83. (<https://doi.org/10.1016/j.jgar.2023.08.016>).

³⁹ ALI et al.: 8. o.



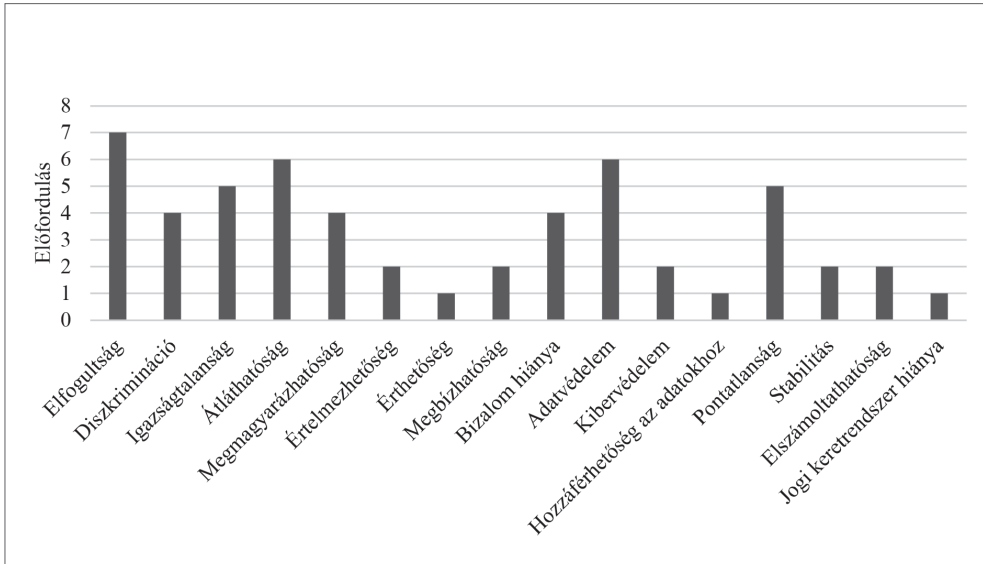
2. ábra. Kockázatok előfordulása 45 szakirodalmi műben



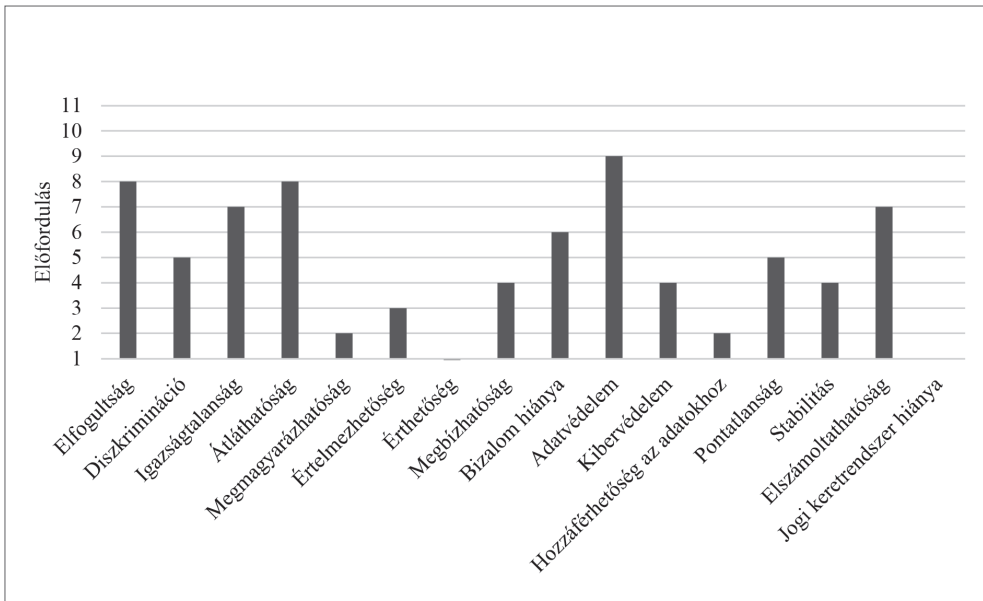
3. ábra. Megjelent kockázatok 2020-ban (7 szakirodalmi mű)

Láthatunk tehát egy tendenciát, hogyan változnak azok a kérdések, amelyek a tudományos szakembereket leginkább foglalkoztatják az MI-alapú döntéshozó és döntéstámogató rendszerekkel kapcsolatban. Az áttekintett szakirodalmakban azonban sokszor minimális mélységgel foglalkoznak az egyes kockázati elemekkel, így azok jól körülhatárolt elemzése a legtöbb esetben elmarad. Definíciós szinten pedig majdnem teljesen mellőzik a kockázatok meghatározását. Kérdésként merülhet fel, hogy

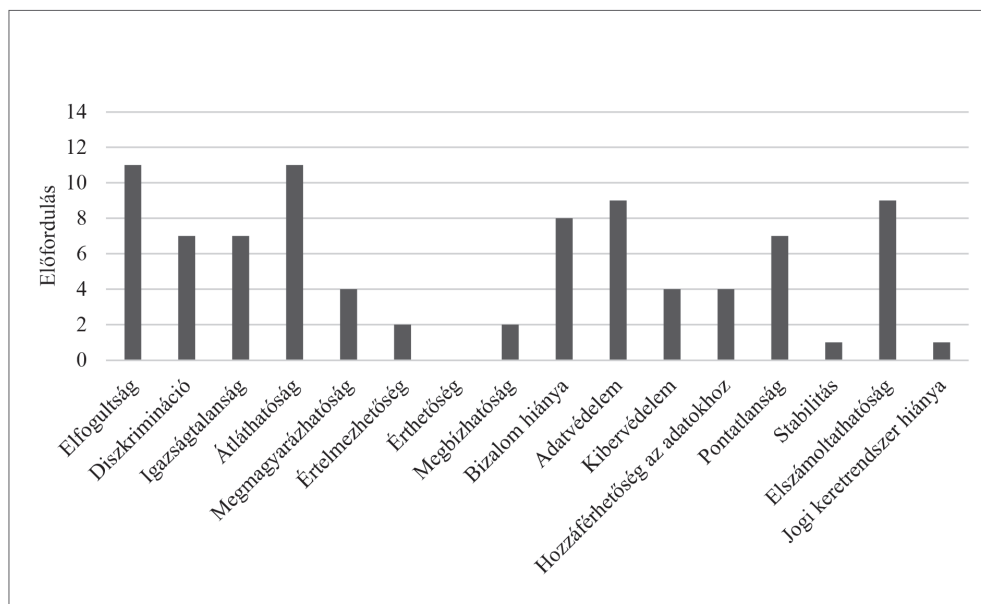
vajon a tudományos szakemberek által vizsgált kockázati elemek párhuzamban vannak-e a piaci szereplők által lényegesnek tartott területekkel. Ennek a kérdésnek a megválaszolására a következő cím alatt teszek kísérletet.



4. ábra. Megjelent kockázatok 2021-ben (8 szakirodalmi mű)



5. ábra. Megjelent kockázatok 2022-ben (11 szakirodalmi mű)



6. ábra. Megjelent kockázatok 2023-ban (14 szakirodalmi mű)

2.3. Eredmények összehasonlítása

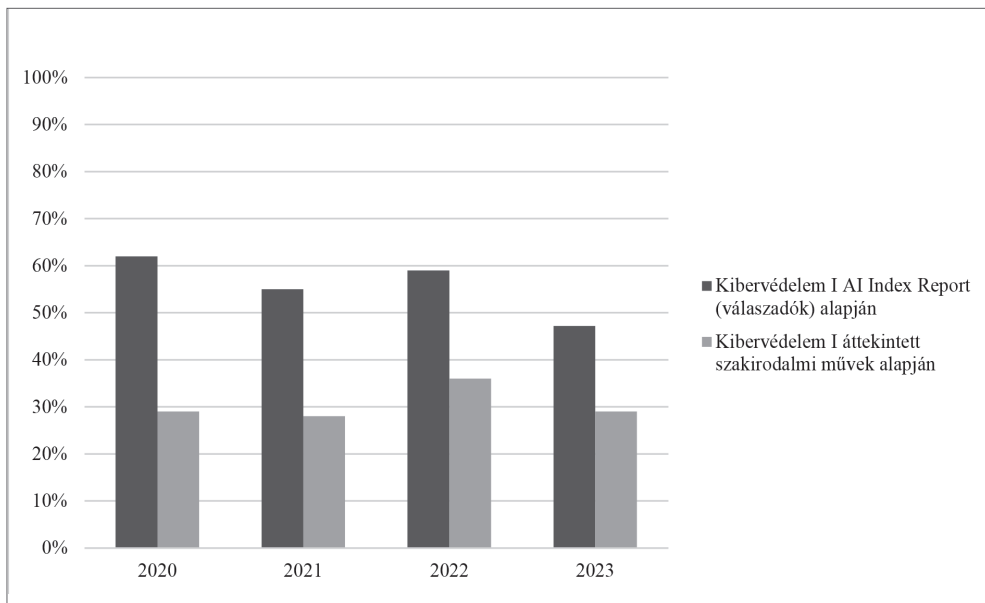
Annak megállapításához, hogy vajon összhangban vannak-e a tudományos szakemberek által leginkább vizsgált kockázati területek a piaci szereplők, azaz a vállalatok, valamint a kis- és középvállalkozások által fontosnak és lényegesnek vélt kockázati területekkel, ahhoz a bevezetőben is hivatkozott, évente megjelenő *AI Index Report*-ot veszem alapul. A 2021-es *AI Index Report*-ban szereplő, a *McKinsey & Company* által végzett kutatás alapján 2019-ben, 1872 megkérdezett cég szerint a kibervédelem a leginkább releváns kockázat (62% szerint) az MI-vel kapcsolatban. Ezt követi a szabályozásnak történő megfelelés (50% szerint), majd a megmagyarázhatóság (45% szerint), a személyes adatok védelme (39% szerint), szervezeti reputáció (35% szerint), munkaerő elmozdulás (34% szerint), méltányosság és igazságosság (beleértve az elfogultságot is, 26% szerint) stb.⁴⁰ A 2022-es évekre ez a tendencia nem változott, így a válaszadók 59%-a szerint a kibervédelem, 45%-a szerint a szabályozásnak történő megfelelés, 40%-a szerint a személyes adatok védelme, 37%-a szerint a megmagyarázhatóság, 32%-a szerint a szervezeti reputáció,

⁴⁰ ZHANG, Daniel–MISHRA, Saurabh–BRYNJOLFFSON, Erik–ETCHEMENDY, John–GANGULI, Deep–GROSZ, Barbara–LYONS, Terah–MANYIKA, James–NIEBLES, Juan Carlos–SELLITTO, Michael–SHOHAM, Yoav–CLARK, Jack–PERRAULT, Raymond: The AI Index 2021 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021.

30%-a szerint a méltányosság és igazságosság és 28%-a szerint pedig a munkaerő-elmozdulás tekinthető releváns kockázatnak.⁴¹

A fentiekől eltérően a 2024-ben közzétett *AI Index Report* alapján a legnagyobb kockázatot több mint 1000 megkérdezett szervezet alapján (20 országból és 19 iparágból) a személyes adatok védelme és az adatok menedzselése jelenti, majd ezt követi az MI-rendszer megbízhatósága, a biztonsága (kibervédelem is), az átláthatóság és az igazságosság.⁴²

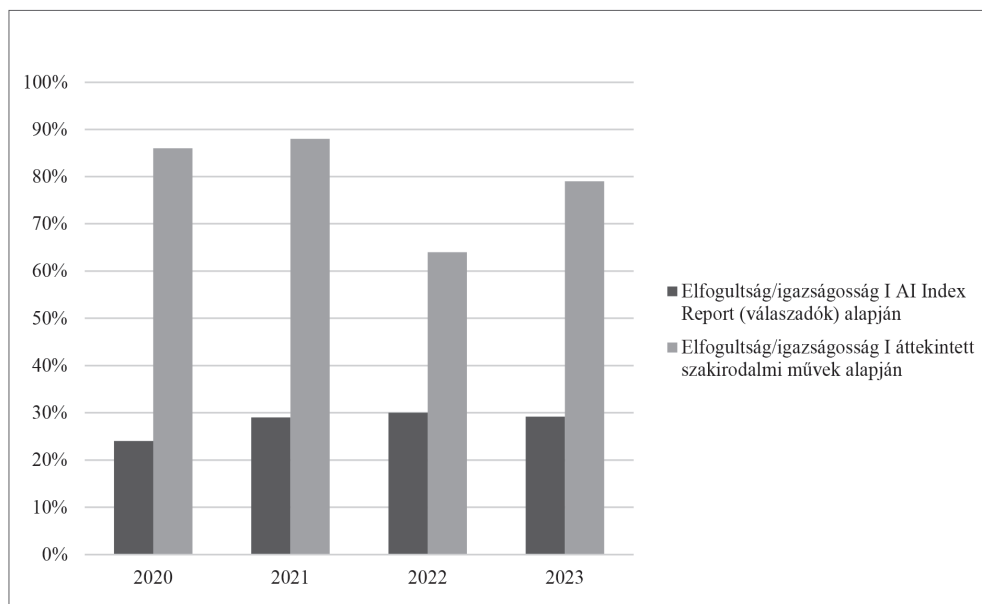
Látható tehát, hogy a tudományos szakemberek által vizsgált kockázatok megjelenése merőben eltérő képet mutat a piaci szervezetek és szereplők által adott válaszoktól. Kivéve ez alól az adatvédelemmel kapcsolatos kockázatokat, hiszen az a 2024-es *AI Index Report*-ban szereplő kutatás alapján is kiemelt helyen szerepel. Mindazonáltal az elfogultságot, amelyet az *AI Index Report* e tekintetben az igazságosság körében értékel, viszonylag kevésbé jelentős kockázatként értékelték (lásd például a 7–9. ábrát).



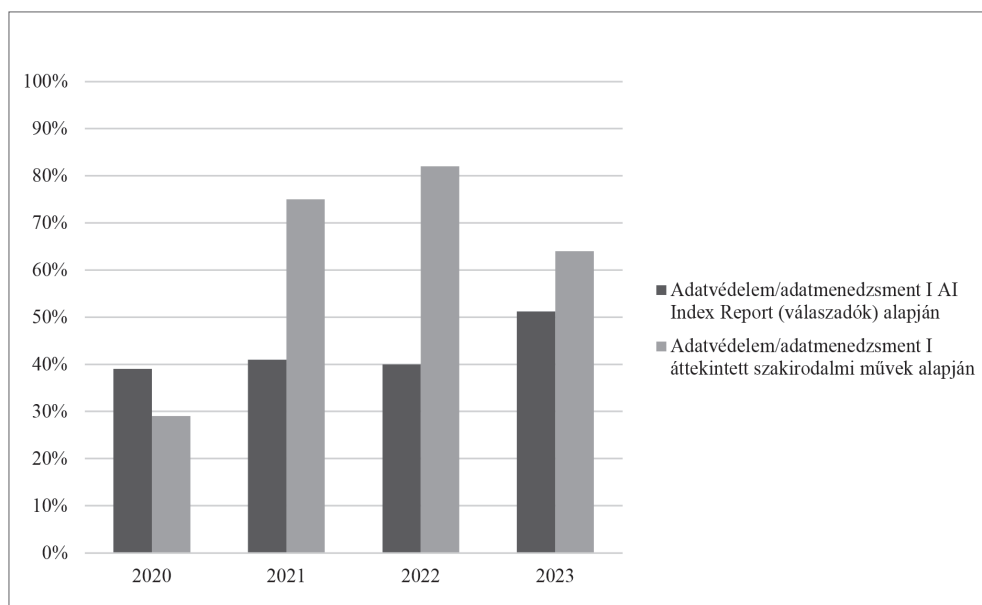
7. ábra. Kibervédelem kockázatként való értékelése

⁴¹ MASLEJ, Nestor–FATTORINI, Loredana–, BRYNJOLFFSON Erik–ETCHEMENDY, John–LIGETT, Katrina–LYONS, Terah–MANYIKA, James–NGO, Helen–NIEBLES, Juan Carlos–PARLI, Vanessa–SHOHAM, Yoav–WALD, Russell–CLARK, Jack–PERRAULT, Raymond: The AI Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.

⁴² MASLEJ–FATTORINI et al.: The AI Index 2024 Annual Report i. m., 166.



8. ábra. Elfogultság/igazságosság kockázatként való értékelése



9. ábra. Adatvédelem/adatmenedzsment kockázatként való értékelése

3. Megoldási javaslatok a kockázatokra

Az áttekintett szakirodalmi művek közül egyes művek a fenti kockázatokra megoldási javaslatokat is prezentálnak, amelyek közül az alábbiakat – jóllehet szubjektív alapon – alkalmasnak vélek a kockázatok mérséklésére és kezelésére. A vállalatok vezetésének – meglátásom szerint – kiemelt figyelmet kell fordítania arra, hogy a korábban említett kockázatokat minimalizálják, vagy legalább megtegyenek mindent annak érdekében, hogy a minimalizálás megvalósuljon. Azon túl, hogy a vállalat vagyoni veszteségét megelőzzék, a vállalati reputáció alóli mentesülést is szolgálhatja, ha megfelelő kockázatmenedzsmenti protokollt alkalmaznak az MI-alapú döntéshozó és döntéstámogató rendszerek vállalati implementációja során.

Ahhoz, hogy ezeket a kockázatok minimalizáljuk, az áttekintett szakirodalom alapján érdemes elsősorban a kontextus oldaláról megvizsgálni a tervezett MI-rendszer bevezetését, azaz egyrészt annak vizsgálatát, hogy területi szempontból milyen szabályozott iparágban szeretnék alkalmazni és milyen speciális szabályozási környezetnek szükséges megfelelni. Ugyanígy szükséges megvizsgálni az MI-rendszer hibás döntéséből származó és egyben várható negatív hatásainak eredményét, azaz hogy milyen hatókörrel, milyen sérülékeny csoportot érintő, vállalati szempontból külső vagy belső hatás várható-e, milyen az emberi jogokra gyakorolt hatása, reverzibilis vagy sem, a várható hatás időtartama, stb.

Az MI-rendszer működtetésének eljárási szempontú vizsgálata során szükséges szemügyre venni azokat a technikai megfontolásokat, amelyek a rendszer komplexitásával kapcsolatosak, vagy például az interoperabilitás fokát. Üzleti megfontolás gyanánt a vállalat kockázattávallási hajlandóságát, az emberi beavatkozás fokát, a kár kezelésére kialakított alternatív protokollokat stb. Technológiai oldaláról szükséges realizálni az algoritmus típusát, történik-e harmadik fél bevonása (például külső informatikai támogatás stb.), milyen szinten érvényesül az átláthatóság, és ezzel együtt a pontosság várható határfoka miként alakul, a rendszer tanulásának gyorsasága stb. Az adatbázis oldaláról érdemes megvizsgálni, hogy az *inputra* adott *output* milyen adatkészletből dolgozik, tartalmaz-e személyes adatokat (esetleg azok különleges kategóriáit), történik-e anonimizálás, mekkora az adatbázis mérete és lefedettsége, stb.⁴³

Amennyiben a menedzsment feltérképezte a fentieket, akkor el lehet gondolkozni azon, hogy milyen típusú MI-alapú rendszert érdemes bevezetni az adott operatív döntéshozatal támogatására. Azaz azt, hogy az emberi és az MI-alapú rendszer általi döntéshozatalnak mekkora teret szánunk. E körben három nagy kategóriát vehetünk számításba. Az emberi autonómia csökkenése szerint haladva elsősorban beszélhetünk az úgynevezett „*human in the loop*” megoldásról, amelyben az MI-alapú rendszer egyszerű támogató rendszerként van jelen. Ilyenkor az MI-rendszer függetlensége alacsony, a döntés meghozatalában a menedzsment teljes bevonásra kerül, így a felelősség is a menedzsmentre hárul. Ennél eggyel autonómabb rendszer az úgynevezett „*human on the loop*” megoldás, ahol a menedzsment autonómiája a jóváhagyásra vagy a döntés elutasítására korlátozódik. E tekintetben az MI-rendszer függetlensége magasabb fokú, viszont a végső döntést még mindig a

⁴³ LEE et al.: i. m.

menedzsment hozza meg. Végül pedig beszélhetünk az úgynevezett „*human out of the loop*” rendszerről, amely már teljesen automatizált, és így a menedzsment nem vesz részt a döntéshozatalban. A fentiek függvényében tehát, ha például az adott döntések meghozatala gyors döntéshozatalt vesz igénybe, a döntés kimenetele várhatóan nem jár emberi jogi jogsértéssel és a döntés meghozatala gyakran ismétlődik, stb., akkor érdemes a *human out of the loop* rendszert kialakítani. Abban az esetben, ahol a döntés meghozatala várhatóan akár emberi jogi jogsértéseket vehet igénybe, a döntés meghozatala többnyire komplex és bonyolult, akkor érdemes a *human in the loop* vagy adott esetben a *human on the loop* rendszert alkalmazni.⁴⁴ Meglátásom szerint a következő években azok a vállalati MI implementációk lesznek sikeresek, amelyeket megelőz egy átfogó kockázatelemzési és kockázatmenedzsmenti munka, annak érdekében, hogy a menedzsment egyfelől csökkentse a potenciális kockázatok bekövetkezési arányát, és ezzel egyúttal a károk előfordulását is kiszorítsa a mindennapi működés során.

Az ilyen típusú MI-alapú döntéshozó és döntéstámogató rendszerek implementálásakor, illetve azok alkalmazása során az Európai Unióban tulajdonképpen két nagy jogszabály által alkotott szabályozási környezetnek kell megfelelni. Egyrészről az automatizált döntéshozatal miatt – amennyiben személyes adatok kezelése is megvalósul – szükséges megfelelni a *természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről* szóló az Európai Parlament és a Tanács 2016/679 rendelete (GDPR) által támasztott speciális szabályoknak. Ilyen speciális szabály, hogy az érintetteket az előzetes tájékoztatás során az automatizált döntéshozatalt alkalmazó adatkezelőnek tájékoztatni kell az automatizált döntéshozatal tényéről, az alkalmazott logikáról, illetve arról, hogy ez milyen jelentőséggel, illetve várható következményekkel bír. Ugyanez a tájékoztatási kötelezettség jelenik meg akkor, amikor az érintett hozzáférési jogával kíván élni.⁴⁵ A GDPR további jogosultságot biztosít az érintetteknek abban az esetben, ha a kizárólag automatizált adatkezelésen (értsd: automatizált döntéshozatalon) alapuló döntés joghatással járna, vagy ehhez hasonlóan jelentős mértékben érintené. Ilyenkor dönthet úgy az érintett, hogy ne terjedjen ki rá ez az adatkezelés. Természetesen e körben is vannak kivételek, amikor ezzel nem élhet az érintett. Ilyen például, ha a döntés az érintett kifejezett hozzájárulásán alapul, vagy az érintett és az adatkezelő közötti szerződés megkötése vagy teljesítése érdekében szükséges, stb. Azonban az érintettnek ilyenkor is van joga ahhoz, hogy az adatkezelő részéről emberi beavatkozást kérjen, álláspontját kifejezze, illetve a döntéssel szemben kifogást nyújtson be.⁴⁶ Kérdésként merülhet fel, hogy az automatizált döntéshozó és döntéstámogató rendszer fentiekben megjelölt három típusa közül melyikre terjed

⁴⁴ IVANOV, S. H.: Automated Decision-Making. *Foresight*, 2023/1, 4–19. (<https://doi.org/10.1108/FS-09-2021-0183>).

⁴⁵ Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről HL L 119., 2016.5.4, p. 1–88. 13. és 15. cikk.

⁴⁶ Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről HL L 119., 2016.5.4, p. 1–88. 22. cikk.

ki ez a speciális szabály. Első ránézésre nyilvánvalóan a *human out of the loop* megvalósulási formája az, amelyre alkalmazni szükséges a GDPR e speciális szabályait. Ezzel szemben az Európai Unió Bírósága a C-634/21. számú ügyében arra a következtetésre jutott, hogy az is automatizált egyedi döntésnek minősül, ha az adatkezelő úgy dönt szerződéskötésről az ügyfele vonatkozásában, hogy egy harmadik szolgáltatót vesz igénybe, annak érdekében, hogy automatizált döntéshozatallal állapítsa meg a szerződéskötés pozitív vagy negatív kimenetelét, és ezt a döntést az adatkezelő automatikusan át is veszi.⁴⁷ Ebből kifolyólag akár a *human in the loop* esete is kimerítheti a GDPR által szabályozott automatikus döntéshozatal megvalósulását. Így a vállalati integráció során erre is kiemelt figyelmet kell fordítani, hiszen érdemes olyan eljárásrendeket kialakítani, amely alapján nem fog a GDPR szerinti automatizált döntéshozatalnak minősülni az adott *human in/on the loop* alapú rendszer által hozott döntés.

Másrészről pedig a *mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról szóló, az Európai Parlament és a Tanács 2024/1689 rendelete* (MI rendelet) adja a megfelelési keretet. A vállalatok, valamint a kis- és középvállalkozások vezetői szintű döntését támogató MI-alapú döntéshozó és döntéstámogató rendszerek esetében eltérő megfelelési kritériumot támaszt az MI rendelet, elsődlegesen attól függően, hogy milyen kategóriába esik maga a rendszer, másodlagosan pedig, hogy milyen minőségben van jelen az adott szervezet (például szolgáltató vagy felhasználó). Példának okáért, ha a munkavállalók teljesítményének nyomon követésével kapcsolatos döntéstámogató rendszert alkalmaznak, amely párosul profilalkotással, akkor az magas kockázatú MI-rendszernek fog minősülni.⁴⁸ Ebben az esetben olyan követelményeknek kell megfelelni, mint például kockázatkezelési rendszer bevezetése, megfelelő adatkormányzás kialakítása, műszaki dokumentáció, nyilvántartás, átláthatóság és az alkalmazóknak nyújtott tájékoztatást stb. Az előbbi követelményeknek való megfelelés azonban csak 2027. augusztus 2. napjától áll fenn,⁴⁹ mindazonáltal érdemes már most ennek megfelelően kialakítani az MI-rendszer vállalati vagy kis- és középvállalkozói integrációját. Megjegyzendő azonban, hogy az előbbi követelményeken túl számos más megfelelési kritériumot is támaszt az MI rendelet, így az adott rendszerre szabva érdemes azonosítani a jogszabályi követelményeket, illetve keretrendszert.

⁴⁷ C-634/21. sz., OQ kontra Land Hessen; a SCHUFA Holding AG részvételével ügyben 2023. december 7-én hozott ítélet (ECLI:EU:C:2023:957).

⁴⁸ Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról HL L, 2024/1689, 2024.7.12, 6. cikk (3) bekezdés.

⁴⁹ Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról HL L, 2024/1689, 2024.7.12, 2. szakasz és a 113. cikk c) pontja.

4. Következtetés és javaslat további kutatásra

A kutatás eredményéből arra a következtetésre jutok, hogy a piaci szereplők számára, azaz a vállalatok, valamint a kis- és középvállalkozások számára nagyobb szükség lenne az MI-alapú rendszerek kibervédelme, adatmenedzsmentje és megbízhatósága szempontjából vizsgálni a kockázatokat, mint az elfogultságra és a fentiekben ismertetett egyéb kockázatra helyezni a hangsúlyt. Továbbá nagy igény lenne az egyes kockázatok mélyebb, adott esetben szektorális vizsgálatára, amely hozzájárulhatna a piaci szereplők kockázatkezeléséhez. Az MI-alapú rendszerekkel kapcsolatos kockázatok közül hiányzik a környezetvédelem. Továbbá hiányzik az ESG (környezet, társadalom, irányítás) szempontú vizsgálata is az ilyen típusú MI-alapú rendszereknek. A fentiek alapján további vizsgálati területként jelölöm meg a munkaerőre gyakorolt hatását ezeknek a támogató rendszereknek. Továbbá célszerűnek vélem a *Massachusetts Institute of Technology* kutatói által 2024 augusztusában közzétett MI-vel kapcsolatos kockázatokat tömörítő adatbázisát⁵⁰ olyan szempontból áttekinteni, amelynek fókuszában a vállalatok, valamint a kis- és középvállalkozások vezetői által alkalmazott MI-alapú döntéshozó és döntéstámogató rendszerek vannak.

5. Limitáció

A jelen kutatással kapcsolatban a következő limitációkat szándékozom kiemelni. A szakcikkek sok helyen nem definiálják az adott kockázatot, azaz azt, hogy pontosan mit is értenek alatta, így előfordul, hogy egyazon kockázatot két különálló néven is azonosítanak. Tekintettel arra, hogy óriási mennyiségű szakirodalom jelenik meg havonta az MI-vel kapcsolatban, így a szakirodalmi áttekintés mint műfaj csak egy pillanatnyi képet tud adni az MI-kockázatokkal kapcsolatos szakirodalomról. Az érintett terület társadalomtudományi szempontú vizsgálatára nem feltétlenül alkalmas a szakirodalmi áttekintés, hiszen ki van téve a szerző szubjektumának.

⁵⁰ SLATTERY, Peter—SAERY, Alexander K.—GRUNDY, Emily A. C.—GRAHAM, Jess—NOETEL, Michael—UUK, Risto—DAO, James—POUR, Soroush—CASPER, Stephen—THOMPSON, Neil: A Systematic Evidence Review and Common Frame of Reference for the Risks from Artificial Intelligence. 2024. (<http://doi.org/10.13140/RG.2.2.28850.00968>).

Függelék. A vizsgált szakirodalmi művek áttekintése

Szakirodalmi mű címe	Vizsgálat területe	Megjelenés dátuma
<i>Are physicians ready for precision anti-biotic prescribing? A qualitative analysis of the acceptance of artificial intelligence-enabled clinical decision support systems in India and Singapore</i>	Egészségügy területén alkalmazott MI-alapú döntéstámogató rendszerek	2023
<i>A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective</i>	Társadalmi problémák az MI döntéshozatali folyamatainak alkalmazásával kapcsolatban	2020
<i>Adjudication of Artificial Intelligence and Automated Decision-Making Cases in Europe and the USA</i>	Jogesetek az MI-vel és az automatizált döntéshozattal kapcsolatban	2023
<i>Towards artificial intelligence at scale in the chemical industry</i>	MI alkalmazása a vegyiparban	2022
<i>AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind</i>	Az MI magyarázhatósági problémájának értelmezése és etikai jelentőségének kiemelése	2021
<i>Analysis of the adoption of emergent technologies for risk management in the era of digital manufacturing</i>	Big data, az MI, a felhőalapú számítástechnika és a blokklánc alkalmazása a kockázatkezelésben	2022
<i>Artificial intelligence technology in surgery</i>	MI jelentősége a sebészetben	2020
<i>Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts</i>	A hibrid döntéshozatali rendszerek különböző formái	2022
<i>Artificial Agents in Corporate Boardrooms</i>	MI használata a vállalati igazgatótanácsokban	2020
<i>Artificial Intelligence and Corporate Decisions: Fantasy, Reality or Destiny</i>	MI szerepe az igazgatótanácsokban	2022
<i>Artificial Intelligence in Health Professions Regulation: An Exploratory Qualitative Study of Nurse Regulators in Three Jurisdictions</i>	MI technológiák lehetséges szerepéről és értékéről a szabályozási gyakorlatban	2023
<i>Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications.</i>	MI technológiák döntéshozatali képességeiről	2019
<i>Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making</i>	Keretrendszer ajánlása az MI hatásainak megértéséhez, a gépi tanulás bevezetésének három fő szakaszára összpontosítva: az adat-, a modell- és az alkalmazási szakaszra	2018

Szakirodalmi mű címe	Vizsgálat területe	Megjelenés dátuma
<i>Artificial intelligence in clinical decision-making: Rethinking liability</i>	A klinikusok és szoftverfejlesztő cégek ellen indított „gondatlansági” keresetek lehetséges kimenetelei az emberi klinikai felügyelet mellett alkalmazott MI-vel működő rendszerek használatával kapcsolatban	2020
<i>Automated Decision-Making</i>	MI-vel kapcsolatos döntéshozatali megközelítések elemzése: „human in the loop”, „human on the loop”, „human out of the loop”	2023
<i>Automating the Risk of Bias</i>	A nemek fokozottabb bevonása az MI technológiák fejlesztésébe, és ennek hatásai	2019
<i>Behind the mask: a critical perspective on the ethical, moral, and legal implications of AI in ophthalmology</i>	Az MI szemészetben és más, az orvostudományhoz kapcsolódó területeken történő alkalmazásának veszélyeiről, ellentmondásos aspektusairól és következményeiről	2023
<i>Beyond the “Black Box”</i>	Átláthatóság és a magyarázhatóság vizsgálata	2021
<i>Challenges in mapping European rare disease databases, relevant for ML-based screening technologies in terms of organizational, FAIR and legal principles: scoping review</i>	Ritka betegségek európai adatbázisainak feltérképezése során felmerülő legfontosabb kihívások azonosítása, amelyek a szervezési, FAIR- és jogi elvek szempontjából relevánsak a géptanulás-alapú szűrési technológiák szempontjából	2023
<i>An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research</i>	Vállalati kormányzás és az MI metszéspontja	2022
<i>Comparing minds and machines: implications for financial stability</i>	Az emberi vagy a mesterséges intelligencia segíti-e jobban a stabil pénzügyi rendszert? – Az emberi és a mesterséges intelligencia döntéshozatali magatartásának aspektusai	2021
<i>Defining the scope of AI ADM system risk assessment</i>	Az MI-ben rejlő definíciók kétértelműségét és annak az automatizált döntéshozattal való kapcsolatát, valamint a rendszer kockázatának szervezeti megértésére gyakorolt lehetséges következményeit vizsgálja	2022
<i>Digital Transformation and Artificial Intelligence Applied to Business: Legal Regulations, Economic Impact and Perspective</i>	Az MI és a digitális transzformáció hatása az üzleti életre	2021
<i>Disappearing Authorship Ethical Protection of AI-Generated News from the Perspective of Copyright and Other Laws</i>	Bemutatja, hogy a jelenlegi cseh jogszabályi környezet nem ösztönzi az elszámoltathatóság, felelősség és átláthatóság megvalósítását	2022

Szakirodalmi mű címe	Vizsgálat területe	Megjelenés dátuma
<i>Distributed Ledgers, Artificial Intelligence and the Purpose of the Corporation</i>	Feltörekvő technológiák hatása mind pozitív, mind normatív szempontból, arra összpontosítva, hogy ezek a fejlemények hogyan befolyásolhatják a tőzsdén jegyzett társaságokkal összefüggésben a vállalati célról folytatott vitákat	2020
<i>Elements of Roboethics</i>	A robotika etikai, jogi és társadalmi vonatkozásai, különös tekintettel a fejlett robotikai alkalmazásokra	2021
<i>Ethical and legal implications of implementing risk algorithms for early detection and screening for oesophageal cancer, now and in the future</i>	A nyelőcsőrák kockázati előre jelző eszközének az alapellátásban való alkalmazásával kapcsolatos etikai és jogi kérdések azonosítása	2023
<i>Ethical considerations on artificial intelligence in dentistry: A framework and checklist</i>	Egy keretrendszer és egy ellenőrző lista nyújtása a fogászatban alkalmazott MI-alkalmazások ilyen szempontból történő értékeléséhez	2023
<i>Ethical machine decisions and the input-selection problem</i>	A ténybeli bizonytalanságnak az erkölcsi döntéshozatalban betöltött szerepe	2021
<i>Ethically Responsible Machine Learning in Fintech</i>	A fintech területén felmerülő etikai kihívások, különösen az elfogultság, a diszkrimináció, a differenciált árképzés, az összeférhetetlenség és az adatvédelem	2016
<i>Explainable AI (XAI) in Rules as Code (RaC): The DataLex approach</i>	A „Rules as Code” megvalósításában a megmagyarázhatóság és az átláthatóság alapvető szükségességét vizsgálja, és megvizsgálja e követelmény hasonlóságát a „magyarázható MI” koncepciójával	2023
<i>Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence</i>	Értékelési módszereket, a rendelkezésre álló eszközöket, a „megmagyarázható MI-adatkészleteket” és más kapcsolódó szempontokat vizsgál	2023
<i>Disability Discrimination in the Digital Realm: How the ICRPD Applies to Artificial Intelligence Decision-Making Processes and Helps in Determining the State of International Human Rights Law</i>	Az MI döntéshozatali folyamatainak a fogyatékkal élő személyekre gyakorolt hatásai	2023
<i>Impact of Artificial Intelligence on Corporate Board Diversity Policies and Regulations</i>	Az MI lehetséges hatásai a vállalati igazgatótanácsok sokszínűségi politikájára és szabályozására	2022
<i>Negligence and AI's Human Users</i>	Az MI egyedi jellegéből származó négy komplikációt vizsgál a gondatlansággal kapcsolatban	2020

Szakirodalmi mű címe	Vizsgálat területe	Megjelenés dátuma
<i>Recommendations for ethical and responsible use of artificial intelligence in digital agriculture</i>	Az MI mezőgazdaságban való alkalmazásának etikai kihívásai, beleértve a méltányosságot, az átláthatóságot, az elszámoltathatóságot, a fenntarthatóságot, a magánélet védelmét és a stabilitást	2022
<i>Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine</i>	Az MI-nek az egészségügyre gyakorolt hatására, valamint az etikus és kiegyensúlyozott beépítés fontosságára összpontosít	2023
<i>Smart Business and the Social Value of AI</i>	Az MI haszon és költség elemzése egy B2B vállalat, valamint annak belső, külső és társadalmi érdekelt felei számára	2022
<i>The Impact of Artificial Intelligence and Digital Economy on Vietnam's Legal System</i>	Az MI és a digitális átalakulás beépülése a különböző alkalmazásokba, és azok szabályozásai	2022
<i>The Integration and Reliance on Technology to Enhance the Independence and Accountability of Company Directors in South Africa</i>	A technológia integrációját és a technológiára való támaszkodást kívánja megvitatni a vállalatirányítási elvek javítása érdekében a fejlődő országokban, például Dél-Afrikában	2021
<i>The perils and promises of generative artificial intelligence in neurointerventional surgery</i>	A generatív MI veszélyeit és ígéreteit vizsgálja a neurointervenciók sebészetben	2023
<i>Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance</i>	Az MI folyamatos fejlődése és adaptációja milyen hatással van a vállalatirányítás gyakorlatára	2020
<i>Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns</i>	A GDPR-alapú átláthatóság követelményének jelentősége az MI és az automatizált döntéshozatali rendszerek esetében	2019
<i>Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?</i>	A GDPR személyes adatok különleges kategóriáira vonatkozó szabályai akadályozzák-e a mesterségesintelligencia-alapú megkülönböztetés megelőzését	2023
<i>Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach</i>	Az algoritmikus döntéshozatalt és a magyarázathoz való jogot vizsgálja	2021