Papp Renáta[1]

A mesterséges intelligencia etikai kihívásai[2]

Absztrakt

Jelen tanulmány, a mestersége intelligencia technológiai fejlődése és alkalmazása során felmerülő etikai kérdéseket és dilemmákat elemzi. A mesterséges intelligencia (MI) egyre nagyobb szerepet tölt be a társadalmi és gazdasági folyamatokban, azonban számos kihívást jelent az adatvédelem, magánélet, döntéshozatali felelősség és igazságosság terén. A tanulmány kiemeli, hogy az MI rendszerek nagy mennyiségű adatot dolgoznak fel, és ezzel növelik az egyének magánéletének sérülékenységét. Az autonóm MI rendszerek, mint az önvezető járművek vagy diagnosztikai eszközök, új felelősségi kérdéseket vetnek fel, különösen, ha döntéseiket minimális emberi beavatkozással hozzák.

Az MI rendszerek döntéshozatala és algoritmusai hajlamosak lehetnek az előítéletek átvételére, ami diszkriminatív eredményekhez vezethet. Az ilyen döntések igazságossága csak akkor biztosítható, ha gondosan fejlesztik és folyamatosan ellenőrzik őket. Az átláthatóság és az elszámoltathatóság létfontosságúak ahhoz, hogy az MI alkalmazások iránti bizalom megmaradjon, hiszen ezek biztosítják a döntések érthetőségét és követhetőségét.

A tanulmány az adatvédelem etikai kereteit vázolja fel, különös tekintettel a GDPR-ra és egyéb nemzetközi iránymutatásokra, amelyek az MI rendszerek etikus és átlátható működését támogatják. Az olyan technológiák, mint a homomorf titkosítás és a federált tanulás, az adatbiztonság erősítésére szolgálnak, miközben csökkentik az egyének magánéletére gyakorolt potenciális negatív hatásokat.

A következtetések arra mutatnak, hogy az MI technológiák felelős fejlesztése multidiszciplináris megközelítést igényel, amely figyelembe veszi az etikai, jogi és társadalmi szempontokat. Az etikus és társadalmilag hasznos MI-alkalmazások csak olyan szabályozási keretek között érhetők el, amelyek biztosítják az emberi jogok tiszteletben tartását, az adatvédelem következetes érvényesítését, valamint a társadalmi egyenlőség és méltányosság megőrzését.

Kulcsszavak: adatvédelem, döntéshozatal, diszkrimináció

Abstract

This study examines the ethical issues and dilemmas that arise with the technological development and application of artificial intelligence (AI). As AI increasingly influences social and economic processes, it presents numerous challenges in terms of data protection, privacy, accountability, and justice. The study highlights that AI systems process large amounts of data, increasing the vulnerability of individual privacy. Autonomous AI systems, such as self-driving vehicles or diagnostic tools, raise new questions of responsibility, especially when decisions are made with minimal human intervention.

AI systems' decision-making and algorithms may be prone to adopting biases, which can lead to discriminatory outcomes. The fairness of such decisions can only be ensured if the

[1] Bírósági fogalmazó (Budapest Környéki Törvényszék), PhD-hallgató (Károli Gáspár Református Egyetem Állam- és Jogtudományi Doktori Iskolája)
[2] DOI szám: 10.59558/jesz.2015.1.51

algorithms are carefully developed and continuously monitored. Transparency and accountability are crucial for maintaining trust in AI applications, as these ensure the understandability and traceability of decisions.

The study outlines the ethical framework for data protection, with particular emphasis on the GDPR and other international guidelines that support the ethical and transparent operation of AI systems. Technologies such as homomorphic encryption and federated learning serve to strengthen data security while minimizing potential negative impacts on individual privacy.

The conclusions suggest that the responsible development of AI technologies requires a multidisciplinary approach that considers ethical, legal, and social aspects. Ethical and socially beneficial AI applications can only be realized within regulatory frameworks that ensure respect for human rights, consistent data protection, and the preservation of social equality and fairness.

Keywords: artificial intelligence, ethical challenges, decision-making, discrimination

Bevezetés

A mesterséges intelligencia (továbbiakban: MI vagy AI – az angol artificial intelligence-ből) egy olyan számítógépes program, amely előre meghatározott algoritmusokat hajt végre. Képes különféle feladatok elvégzésére, nagyméretű adathalmazok elemzésére és mintázatok felismerésére. Ezek alapján következtetéseket von le, jövőbeli eredményeket jósol, és a rendelkezésre álló adatok alapján megalapozott döntéseket hoz.[3] A mesterséges intelligencia fejlesztésével és széles körű alkalmazásával járó etikai kihívások rendkívüli jelentőséggel bírnak, mivel ezen technológiák hosszú távon alapvető társadalmi és gazdasági hatásokat idézhetnek elő. Az MI etikai kérdései között különös figyelmet érdemel az adatvédelem és a magánélet védelme, az etikus és pártatlan döntéshozatal, valamint az emberi méltóság és jogok tiszteletben tartása. Az adatvédelem és a magánszféra védelmének kérdése azért kerül előtérbe, mert az MI rendszerek általában nagy mennyiségű, személyes és érzékeny adatot dolgoznak fel, ami – a felhasználás céljától és módjától függően – potenciálisan veszélyeztetheti az egyén magánéletét és személyes autonómiáját. Az ilyen adatfeldolgozási gyakorlatok különösen veszélyesek lehetnek az érzékeny adatok kezelésénél, ahol az átláthatóság és a felhasználók tájékoztatása elengedhetetlen.[4]

Az MI etikus döntéshozatali kihívásai is kiemelkedő jelentőségűek, mivel ezen rendszerek hajlamosak lehetnek az adatokban rejlő előítéletek és torzítások átvételére, ami diszkriminatív eredményeket szülhet, ha a tervezés és tanítás során nem megfelelő módon kezelik ezeket a kockázatokat.[5] Ennek következtében az MI döntéshozatalának megbízhatósága és igazságossága csak akkor biztosítható, ha az algoritmusokat gondosan fejlesztik és rendszeresen ellenőrzik. Emellett az emberi méltóság és jogok védelme

[3] Glenn, Gordon: The Use of Artificial Intelligence in the Legal Profession https://www.lexisnexis.com/community/insights/legal/practical-guidance-journal/b/pa/posts/the-use-of-artificial-intelligence-in-the-legal-profession?srsltid=AfmBOorWKjTDsoYam uuR2leSQfzwQqGoKodMTdevfkX2_AwrHllspz5 (2024. október 21.)

[4] Centre for Information Policy Leadership (CIPL) (2020): Artificial Intelligence and Data Protection How the GDPR Regulates AI. 3. o., https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (2024.október 21.)

[5] Corbett-Davies, Sam – Pierson, Emma – Feller, Avi – Goel, Sharad – Huq, Aziz: Algorithmic decision making and the cost of fairness. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13-17 August 2017. 797-806. o.

kulcsfontosságú szempont minden MI-alkalmazásban, hogy biztosítsák az emberi értékek és etikai normák tiszteletben tartását minden olyan területen, ahol az MI rendszerek befolyással bírnak.[6]

Ezen etikai kihívások kezelése átfogó és multidiszciplináris megközelítést igényel a mesterséges intelligencia fejlesztésében és alkalmazásában, hogy biztosítható legyen az MI technológia előnyeinek fenntartható és felelős kiaknázása, miközben minimalizálják annak potenciális negatív hatásait.

I. Felelősség az autonóm döntéshozatalban

Az autonóm döntéshozatal és a felelősség kérdései a mesterséges intelligencia fejlesztésének és alkalmazásának egyik legösszetettebb etikai területét jelentik. Az autonóm döntéshozatalra képes MI-rendszerek – különösen azok, amelyek kritikus döntéseket hoznak az emberi beavatkozás minimális szükségessége mellett – alapjaiban változtatják meg a felelősség értelmezését és alkalmazását. Ilyen rendszerek például az önvezető járművek, az egészségügyi diagnosztikai eszközök, vagy a pénzügyi és jogi döntéshozatali rendszerek,[7] amelyek gyakran emberi felügyelet nélkül működnek, és döntéseiket automatizált algoritmusok irányítják. Az ilyen rendszerek működésében felmerül a kérdés: ki vállalja a felelősséget, ha az autonóm döntéshozatal során hibás döntések születnek vagy károk keletkeznek?

Az átláthatóság egy komplex fogalom, amely több aspektust foglal magában, mint például a magyarázhatóság, értelmezhetőség, nyitottság és hozzáférhetőség. Különböző tudományágak eltérő nézőpontokat hangsúlyoznak: a közgazdászok az optimális piacok, a politológusok a politikai részvétel, a jogtudósok az adminisztratív törvényesség előfeltételének tekintik. Az átláthatóság három fő perspektívában értelmezhető: mint erény, kapcsolat és rendszer. Az erény nézőpontjában az átláthatóság normatív érték, amely meghatározza a nyitottságot. A kapcsolati nézőpont a befogadó és a szereplő közötti viszonyra összpontosít, míg a rendszerszemlélet az intézményi környezetbe való beágyazottságot hangsúlyozza. E három nézőpont integrálása elengedhetetlen az átláthatóság teljes megértéséhez.[8]

Az autonóm döntéshozatal felelősségi kérdései különösen bonyolultak az olyan helyzetekben, ahol az MI döntései közvetlen hatással vannak emberek életére vagy biztonságára. Ha egy autonóm rendszer döntése káros kimenetelt eredményez, a felelősség kiosztása több lehetséges szereplő – mint a fejlesztő, a gyártó, az üzemeltető vagy a felhasználó – között is megoszlik, ami megnehezíti az egyértelmű felelősségi viszonyok megállapítását. Az ilyen esetekben a "felelősségi szakadék"[9] fogalma merül fel, amely arra utal, hogy az autonóm rendszerek által hozott döntésekért senki nem vállalja közvetlenül a felelősséget, vagy legalábbis nem egyértelmű, hogy ki lenne felelős a következményekért. Ez a szakadék különösen problémás, mivel az MI önálló döntéshozatali képessége miatt a

[6] UNESCO (2021): Recommendation on the Ethics of Artificial Intelligence: Respect, protection and promotion of human rights and fundamental freedoms and human dignity. 18. o. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (2024. október 17.)

[7] Mehdi, Dastani – Vahid, Yazdanpanahof: Responsibility of AI Systems, In AI & Soc, 2023. 38, 843. o. 1. bek.

[8] Felzmann, Heike – Fosch-Villaronga, Eduard – Lutz, Christoph – Tamó-Larrieux, Aurelia: Towards Transparency by Design for Artificial Intelligence. In Sci Eng Ethics No. 26, 2020, 3335-3336. o.

[9] A kérdést számos szerző tárgyalta, a legújabbak közül néhány: (Danaher, John: Tragic Choices and the Virtue of Techno-Responsibility Gaps. In Philosophy & Technology, No. 2, 2022, 1-26. o.; Chengeta, Thompson: Accountability gap: Autonomous weapon systems and modes of responsibility in international law. In Denver Journal of International Law and Policy, No. 1, 2016.; Crawford, C.: Individual and Collective Moral Responsibility for Systemic Military Atrocity. In Journal of Political Philosophy, No. 2, 2007, 187-212. o.

hagyományos felelősségre vonási rendszerek nem mindig alkalmazhatók megfelelően.

Az átláthatóság kulcsfontosságú az erőforrások hatékony elosztásában és az elszámoltathatóság biztosításában, mivel csökkenti az információs aszimmetriákat. Az AI rendszerekben különösen fontos a magyarázhatóság és ellenőrizhetőség szempontjából, lehetővé téve a rendszerek működésének nyomon követését és ellenőrzését. Noha az átláthatóság és az elszámoltathatóság szorosan összefügg, nem azonos fogalmak. Az átláthatóság szintén hozzájárulhat a bizalom építéséhez, különösen technológiai környezetben, bár ennek hatásai összetettek és további vizsgálatot igényelnek.[10]

Az autonóm MI rendszerek esetében gyakran javasolt az átláthatóság növelése, hogy a döntési folyamat és annak indoklása a lehető legjobban megérthető és rekonstruálható legyen. Az átláthatóság és a nyomon követhetőség biztosítása érdekében egyes kutatók és szabályozók olyan etikai és technológiai keretrendszerek kialakítását szorgalmazzák, amelyek az MI rendszer fejlesztőire, üzemeltetőire és felhasználóira egyaránt kötelezettségeket rónak. Az autonóm rendszerek felelősségi keretének létrehozása elengedhetetlen ahhoz, hogy az MI rendszerek döntéseit az emberi felhasználók és a társadalom is bizalommal fogadja, elősegítve ezzel a technológia biztonságos és etikus elterjedését.

I.1. Algoritmusok és autonóm rendszerek döntéshozatala

Az algoritmusok által vezérelt döntéshozatal és az autonóm rendszerek működése a mesterséges intelligencia alkalmazásának egyre elterjedtebb formáivá váltak, amelyek mély társadalmi és etikai kihívásokat vetnek fel. Az MI alapú algoritmusok nagy mennyiségű adat feldolgozásával képesek gyors és hatékony döntéseket hozni olyan kritikus területeken, mint a pénzügyek, az egészségügy, a közlekedés vagy épp a bűnüldözés. Az ilyen rendszerek azonban csak annyira pártatlanok és igazságosak, mint azok az adatok, amelyekkel tanították őket; így, ha a tanító adatok torzítottak vagy hiányosak, az algoritmusok hajlamosak továbbvinni és felerősíteni a meglévő előítéleteket és diszkriminatív mintázatokat. Az algoritmusok által vezérelt döntések további hátránya, hogy sok esetben hiányzik belőlük az emberi ítélőképesség és empátia, amelyek a komplex és egyedi helyzetek kezelésében elengedhetetlenek. Az ilyen döntési mechanizmusok etikai kihívásai különösen érzékeny területeken, például az egészségügyben és a jogrendszerben jelentkeznek, ahol a személyes körülmények és az egyéni méltányosság alapvető fontosságú.

Az MI autonóm működésének és az algoritmusok döntéshozatalának elterjedésével egyre több kérdés merül fel az emberi beavatkozás és felelősség kapcsán. Az algoritmikus döntéshozó rendszerek (Algorithmic Decision Systems (ADS)[11] használata különböző helyzetekben és különböző hatásokkal járhatnak. Az ADS-ek számos területen használatosak, például az egészségügyben (ahol egy hibás algoritmus komoly károkat okozhat a betegek kezelésében), igazságszolgáltatásban, banki és biztosítási szektorban, e-kereskedelemben, valamint autonóm rendszerekben, mint például önvezető autókban, ahol egyre több olyan döntést képesek önállóan meghozni, amelyek hagyományosan emberi felügyeletet igényeltek volna. Bár ezek a rendszerek jelentős előnyökkel járhatnak, mint például jobb döntéshozatal és hatékonyság, kockázatokat is hordoznak, például diszkrimináció, manipuláció vagy adatvédelmi problémák formájában. Jelentős előnyöket kínálnak az egyének, a magánszektor és a közszektor számára, például a hatékonyság, a döntéshozatal javítása és a személyre szabott szolgáltatások terén. Azonban komoly kockázatokat is hordoznak, mint a

[10] Felzmann – Fosch-Villaronga – Lutz – Tamò-Larrieux: i.m. 3336-3339. o.

[11] Algorithmic Decision Systems (ADS): Az ADS általában a döntéshozatali folyamat alapjául szolgáló többfolyamatos algoritmusokon keresztül történő automatizálásra utal, beleértve az adatok gyűjtését és feldolgozását, valamint a döntések végrehajtását csekély emberi beavatkozással vagy anélkül.

diszkrimináció, adatvédelmi problémák és az autonómia elvesztése. A magánszektorban az algoritmusok alkalmazása torzíthatja a piaci versenyt, míg a közszektorban a méltányosság és az átláthatóság hiánya bizalmatlansághoz vezethet. A kockázatok minimalizálása érdekében alapvető fontosságú a szabályozás, amely biztosítja az átláthatóságot, elszámoltathatóságot és adatvédelmet.[12]

Ezért kiemelkedően fontos annak meghatározása, hogy milyen mértékű emberi felügyelet és ellenőrzés szükséges a rendszer működéséhez, illetve ki viseli a felelősséget az MI által hozott döntések következményeiért.[13] Az autonóm rendszerek döntéshozatali képességei kapcsán felmerül a "felelősségi szakadék" problémája, amely arra utal, hogy ezek a rendszerek nem rendelhetők egyértelműen felelősség alá, ami zavarokat okozhat a felelősség és az elszámoltathatóság érvényesítésében.

Az etikai és jogi keretrendszerek kialakítása nélkülözhetetlen annak érdekében, hogy a mesterséges intelligencia rendszerek átlátható, felelős és igazságos módon működjenek. A mesterséges intelligencia fejlesztése jelentős előnyöket kínál, de komoly aggodalmakat vet fel a magánélet, elfogultság és átláthatóság terén. Emiatt szükség van jogi keretek kialakítására, amelyek biztosítják az AI-rendszerek elszámoltathatóságát és védik az emberek jogait. A jogi keretek kulcselemei közé tartozik az átláthatóság, az adatvédelem, a méltányosság és elfogultság kiküszöbölése, az elszámoltathatóság egyértelmű meghatározása, a rendszerek szigorú tesztelése és az etikai megfontolások figyelembevétele.[14] A mesterséges intelligencia fejlődésével számos jogi keret segíti az elszámoltathatóság biztosítását. A GDPR[15] szabályozza a személyes adatok kezelését, az Algorithmic Accountability Act (AAA)[16] az algoritmusok átláthatóságát és méltányosságát célozza, míg az OECD alapelvei[17] és az Európai Bizottság irányelvei az AI etikus és átlátható használatát támogatják. Ezek a jogi keretek az átláthatóság, adatvédelem, méltányosság, elszámoltathatóság, valamint etikai megfontolások figyelembevételével segítenek kezelni az AI-rendszerekkel kapcsolatos etikai és jogi aggályokat. Az olyan nemzetközi szervezetek, mint az Európai Bizottság[18] és az IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, különböző iránymutatásokat dolgoztak ki, amelyek az algoritmusok és autonóm rendszerek átláthatóságát, társadalmi felelősségét, valamint az emberi felügyelet fontosságát hangsúlyozzák. Továbbá az Európai Bizottság „Ethics Guidelines for Trustworthy AI”[19] dokumentuma például részletes útmutatást nyújt az MI rendszerek átláthatóságáról, az előítéletek minimalizálásáról, valamint a társadalmi felelősségvállalásról, külön hangsúlyozva

[12] European Parliament: Understanding algorithmic decision-making: Opportunities and challenges. STOA | Panel for the Future of Science and Technology. Scientific Foresight Unit (STOA) (2019). 3-6., 7-18. o. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf (2024. július 24.)

[13] Ashraf, Afsa (2022): Why Artificial Intelligence Requires Human Intervention. https://www.royalcyber.com/blogs/ai-systems-need-human-intervention/ (2024.07.17.)

[14] Grady, Andersen – MoldStud Research Team: Ethics in Artificial Intelligence Systems Analysis: Ensuring Fairness and Accountability. Legal Frameworks for Ensuring Accountability in Artificial Intelligence Systems. (2024). https://moldstud.com/articles/p-ethics-in-artificial-intelligence-systems-analysis-ensuring-fairness-and-accountability (2024.október 21.)

[15] GDPR: Az európai parlament és a tanács 2016. április 27-i (EU) 2016/679 rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet). https://net.jogtar.hu/jogszabaly?docid=a1600679.eup (2024. október 16.)

[16] S.2892 – Algorithmic Accountability Act of 2023: https://www.congress.gov/bill/118th-congress/senate-bill/2892/text (2024. október 16.)

[17] Az OECD Multinacionális vállalatokra vonatkozó irányelvei (Irányelvek): https://oecdmnkp.hu/hu/iranyelvek (2024. október 17.)

[18] Európai Bizottság hivatalos honlapja: https://commission.europa.eu/index_hu (2024. október 17.)

[19] Ethics guidelines for trustworthy AI (2019). European Comission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2024.07.17.)

az emberi felügyelet és ellenőrzés szükségességét.

A mesterséges intelligencia fejlődése során az algoritmusok által vezérelt döntéshozatal etikai kereteinek kidolgozása és a felelősségi kérdések tisztázása alapvető fontosságú annak érdekében, hogy a társadalmi normáknak megfelelően alkalmazzák az MI rendszereket. Az átláthatóság növelése, az emberi beavatkozás biztosítása, valamint az algoritmusok etikusan tervezett és ellenőrzött működése hozzájárulhat az MI bizalomépítő alkalmazásához, ezáltal biztosítva a technológia fenntartható és társadalmilag hasznos fejlődését.

I.2. Diszkrimináció és igazságosság

Az MI-rendszerek óriási mennyiségű, történeti adatokon alapuló információk felhasználásával kerülnek betanításra, amelyek gyakran magukban hordozzák a társadalmi torzításokat és előítéleteket. Ennek következtében ezek a rendszerek hajlamosak az adatokban rejlő elfogultságokat beépíteni az algoritmusok működésébe, amely nemcsak állandósíthatja, de felerősítheti is az igazságtalan vagy diszkriminatív kimeneteleket olyan kulcsfontosságú területeken, mint a munkaerő-felvétel, a hitelbírálat, a büntető igazságszolgáltatás, valamint a társadalmi erőforrások elosztása.

Például egy vállalat, amely mesterséges intelligencia rendszert használ az állásra jelentkezők előszűrésére az önéletrajzok elemzése alapján, gyakran olyan múltbeli adatokra támaszkodik, amelyek a vállalat eddigi sikeres toborzási gyakorlatát tükrözik. Azonban, ha ezek a korábbi adatok elfogultak – például nemi, faji vagy egyéb társadalmi előítéletek hatását mutatják –, akkor az MI rendszer szinte biztosan átveszi ezeket a mintázatokat, így indirekt módon diszkriminálhatja azon jelölteket, akik eltérnek a vállalat korábbi preferenciáitól.[20]

A vallási intézmények toborzási gyakorlata is jól illusztrálja a diszkrimináció fogalmának összetettségét és az etikai megfontolások változó szerepét. Korábban a vallási intézmények saját kritériumokat alkalmazhattak, például a vallási elkötelezettséget követelhették meg az álláspályázóktól, de az Európai Bíróság ítéletei ezt korlátozták. A bíróság kimondta, hogy a vallási követelmények csak akkor érvényesek, ha azok szorosan kapcsolódnak az adott tevékenységhez és arányosak. Ez a dinamikus súlyozás új megvilágításba helyezi a diszkrimináció kérdését, és megmutatja, hogy különböző társadalmi és jogi kontextusokban eltérő módon értelmezhetik.[21]

Hasonló kihívások merülnek fel a pénzügyi szektorban is, ahol az MI rendszerek hitelminősítési folyamatai során előfordulhat, hogy a múltbeli adatok alapján generált modell bizonyos társadalmi csoportokat hátrányosan érint, ezzel hosszú távon mélyítve a gazdasági egyenlőtlenségeket.

A jelenlegi büntetőjogi igazságszolgáltatási rendszerben az emberi döntéshozatal jelentős szerepet játszik, ám ez számos problémát hordoz magában, különösen az egyenlőség és az igazságosság szempontjából. Amerikai kutatások szerint a döntésekben, például letartóztatás, vádemelés, ítélethozatal és feltételes szabadlábra helyezés során gyakran megjelennek a tudattalan előítéletek és implicit elfogultságok, amelyek aránytalanul sújtják a kisebbségi csoportokat, különösen a feketéket és latinókat.[22] A bírák, ügyészek és rendőrök

[20] The Ethical Considerations of Artificial Intelligence, Capitol, Bias and Discrimination (2023), Capitol Technology University https://www.captechu.edu/blog/ethical-considerations-of-artificial-intelligence (2024. október 16.)

[21] Heinrichs, Bert: Discrimination in the age of artificial intelligence, 147-148. o. In AI & Society, No. 37, 2022, 143-154. o.

[22] Kleinberg, Jon – Ludwig, Jens – Mullainathan, Sendhil – Sunstein, Cass R: Discrimination in the Age of Algorithms, 116. o. In Journal of Legal Analysis, No. 10, 2018, 113-174.

döntéseit gyakran befolyásolják személyes preferenciáik és korlátozott racionalitásuk, ami az egyes csoportok számára hátrányos ítéletekhez vezet.[23] Ezért szükséges az emberi döntéshozatal javítása a kriminalisztikában.

A probléma súlyosságát felismerve, több amerikai szabályozó hatóság az utóbbi időszakban figyelmeztetéseket adott ki, amelyek célja az AI-algoritmusok elfogultságának mérséklése, valamint a vállalatok felelősségre vonása a platformjaikon keresztül megvalósuló diszkriminációért. Ezek az intézkedések hangsúlyozzák a felelősségi körök tisztázásának és az etikai keretrendszerek megalkotásának fontosságát, amelyek biztosítják, hogy az AI rendszerek átlátható módon, az egyenlőség és a tisztesség szempontjait figyelembe véve működjenek, csökkentve az előítéletek kockázatát. A globális szabályozói és kutatói közösség számos iránymutatást dolgozott ki ezen problémák megoldására, amelyeket olyan kulcsfontosságú dokumentumok rögzítenek, mint az Európai Bizottság " Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan"[24] és az IEEE[25] globális irányelvei. Ezek az iránymutatások szorgalmazzák az algoritmusok átláthatóságának és a társadalmi felelősségvállalásnak a növelését, valamint az emberi felügyelet és ellenőrzés biztosítását, amely elengedhetetlen a tisztességes és megbízható MI alkalmazások elterjedéséhez. Az IEEE Szabványügyi Szövetség (IEEE SA)[26] számos autonóm és intelligens rendszerekhez (IEEE 7000)[27] kapcsolódó szabványt és szabványprojektet kínál, amelyek az AI, robotika, kiberfizikai rendszerek és más kapcsolódó technológiák területén alkalmazhatók. Ezek a szabványok különféle területeket ölelnek fel, például a mesterséges intelligencia diagnosztikai interfészeit, a karbantartási rendszerek interoperabilitását, a kiterjesztett valóság tanulási modelleket, valamint a biometrikus adatvédelem és gépi tanulási alkalmazások keretrendszereit. Az IEEE SA szabványai biztosítják az etikus tervezés, az adatvédelem, az átláthatóság és a biztonság követelményeit az autonóm rendszerek, mesterséges intelligencia és digitális rendszerek fejlesztése során.[28]

I.3. Igazságosság elvei az MI alkalmazásában

Az igazságosság elvei kulcsfontosságú szerepet játszanak az MI-rendszerek etikus és elfogadható alkalmazásában, biztosítva, hogy az algoritmusok által hozott döntések

[23] Michelle, Alexander: The New Jim Crow: Mass Incarceration in the Age of Colorblindness New Press, 2012, 312. o. (2024. október 19.) Ez a könyv bemutatja, hogy az igazságszolgáltatásban meglévő előítéletek és diszkrimináció hogyan vezethet rendszerszintű hátrányokhoz bizonyos társadalmi csoportok, különösen a kisebbségek számára.

[24] Lásd bővebben: Mesterséges intelligenciával foglalkozó magas szintű független szakértői csoport (2019): Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan. https://op.europa.eu/hu/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (2024. október 19.)

[25] Lásd: Az Institute of Electrical and Electronics Engineers egy nemzetközi szakmai szervezet, amely elektronikai, villamosmérnöki, valamint számítástechnikai tudományokhoz kapcsolódó szabványok kidolgozásával, oktatással, tudományos konferenciák szervezésével, illetve ezekhez kötődő tanulmányok publikálásával foglalkozik. Institute of Electrical and Electronics Engineers (IEEE) hivatalos honlapja: https://www.ieee.org/ (2024. október 19.)

[26] Lásd: Az Institute of Electrical and Electronics Engineers Standards Association az IEEE-n belüli működési egység, amely globális szabványokat fejleszt számos iparágban.

[27] Az IEEE 7000 nevű etikus szabvány segíthet a vállalatoknak és a fejlesztőknek abban, hogy tisztességes intelligens rendszereket készítsenek. Az IEEE a világ legnagyobb ipari szövetsége, amelynek világszerte több mint 420 000 tagja van. A szervezet úgy döntött, hogy ideje megteremteni az etikus és tisztességes autonóm és intelligens rendszerek kialakításának az alapjait. Az IEEE 7000 szabványt már számos cég tesztelte a gyakorlatban. Forrás: https://m.sg.hu/#gsc.tab=0 (2024. október 19.)

[28] Institute of Electrical and Electronics Engineers Standards Association (IEEE SA): Autonomous and Intelligent Systems (AIS) Standards. https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/ (2024. október 19.)

igazságosak, átláthatóak és diszkriminációmentesek legyenek. Az MI technológiák növekvő hatása az élet számos területére, különösen a munkaerőpiacra, a közszolgáltatásokra és a közbiztonságra, kiemeli annak fontosságát, hogy ezen rendszerek működése a társadalmi értékeket és egyenlőséget szolgálja. Az igazságosság biztosításához számos alapvető elvet szükséges alkalmazni, melyek célja, hogy az MI előnyeit a legszélesebb körben biztosítsák, miközben minimalizálják a technológiából adódó etikai és társadalmi kockázatokat.

Diszkriminációmentesség: A mesterséges intelligencia rendszerek gyakran képesek különféle döntéshozatali folyamatok automatizálására és gyorsítására, de ugyanakkor az egyenlőség és diszkriminációmentesség kockázatát is hordozhatják, mivel az adatok és a modelltervezés torzításai előítéletes eredményekhez vezethetnek. Ezen kihívások kezelésére az Európai Rasszizmus és Intolerancia Elleni Bizottság (European Commission against Racism and Intolerance, ECRI)[29] irányelvei arra ösztönzik az államokat, hogy magas szinten – akár alkotmányos szinten – kötelezzék el magukat az egyenlő bánásmód mellett. Az ajánlások szerint a nemzeti jogszabályoknak világosan kell kezelniük a hátrányos megkülönböztetést, valamint szigorúan büntetniük kell a rasszista cselekedeteket. Az oktatásban is erősíteni kell a kultúrák sokszínűségének megbecsülését és a jogi ismeretek terjesztését, míg a közszolgálatokban a kisebbségek alkalmazását kell támogatni, továbbá biztosítani kell, hogy az állampolgárok egyenlő hozzáférést kapjanak a közszolgáltatásokhoz. Ezen intézkedések célja, hogy elősegítsék az MI rendszerek diszkriminációmentes működését a közszolgáltatásoktól a munkahelyi döntéshozatalig.[30]

Transzparencia és magyarázhatóság: Az MI rendszerek gyakran „fekete dobozként" működnek, vagyis belső folyamatuk és döntéshozatali mechanizmusaik csak korlátozottan érthetők meg, ami megnehezíti az átláthatóságot és a következetes ellenőrzést. Különösen kritikus ez az olyan területeken, mint az egészségügy vagy az autonóm járművek, ahol elengedhetetlen annak tisztázása, hogy miként születnek a döntések, és ki vállalja a felelősséget az esetleges hibákért vagy károkért. Az elszámoltathatóság biztosítása különösen fontos abban az esetben, ha az MI rendszerek működése nem kívánt következményekkel jár, lehetővé téve a szükséges korrekciós intézkedések alkalmazását és a hibák kivizsgálását. A fekete doboz probléma megoldására a kutatók a megmagyarázható mesterséges intelligencia (Explainable AI[31]) fejlesztésére összpontosítanak, amely segíti a modellek átláthatóságát, igazságosságának, pontosságának és lehetséges torzításainak megértését.[32]

Az MI rendszerek átláthatósága kritikus fontosságú az emberi jogok és a tisztességes eljáráshoz való jog biztosítása érdekében. Az átláthatóság hiánya aláássa a döntések legitimitását, és sérti az egyének jogát, hogy megértsék és ellenőrizzék az MI által meghozott döntéseket, különösen akkor, ha azok közvetlenül befolyásolják az életüket, biztonságukat vagy alapvető jogaikat. Az MI döntéseinek megmagyarázhatósága lehetővé teszi, hogy az érintettek megértsék az algoritmusok működésének logikáját, így növelve a bizalmat a rendszerek iránt.

Felelősség és elszámoltathatóság: A mesterséges intelligencia elszámoltathatósága a technológia etikus és jogszerű használatának biztosítására épül, különösen az automatizált döntéshozatal terén. Az elszámoltathatóság keretei a hatalom átruházásával, a hatóságok által

[29] Lásd: Európai rasszizmus és intolerancia-ellenes bizottság az ECRI általános irányelv ajánlása no. 1: harc a rasszizmus, az idegengyűlölet, az antiszemitizmus és az intolerancia ellen., https://rm.coe.int/ecri-general-policy-recommendation-no-1-on-combating-racism-xenophobia/16808b59e9 (2024. október 16.)

[30] Lásd: A rasszizmus és intolerancia elleni európai bizottság 7. Sz. Általános ajánlása: a rasszizmus és a faji megkülönböztetés elleni küzdelem a nemzeti jogalkotásban, https://rm.coe.int/ecri-general-policy-recommendation-no-7-revised-on-national-legislatio/16808b5ab3 (2024. október 16.)

[31] Nicklas, Ankarstad: What is Explainable AI (XAI)? Medium, (2020) https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513 (2024. október 16.)

[32] Capitol Technology University: The Ethical Considerations of Artificial Intelligence, 3. bek. (2023) https://www.captechu.edu/blog/ethical-considerations-of-artificial-intelligence (2024. október 16.)

végzett vizsgálatokkal és a hatalom korlátozásával kapcsolatos feltételekre épülnek. Az AI elszámoltathatósága több szinten is megjelenik: a szabályozás, a jelentési kötelezettségek, a felügyelet és a jogérvényesítés. Az AI rendszereknél fontos kérdés az, hogy ezek az egyedi rendszerek milyen módon felelnek meg etikai és jogi normáknak, és miként működnek együtt az emberi felügyelettel a társadalmi rendszerekben.

A mesterséges intelligencia rendszerek elszámoltathatósága alapvető kérdéseket vet fel, amelyek jogi, etikai és társadalmi szempontból egyaránt jelentőséggel bírnak. Mivel az AI technológia egyre szélesebb körben alkalmazott, a társadalom számára kiemelkedően fontos, hogy világos és átlátható szabályozási kereteket alakítsunk ki annak érdekében, hogy meghatározható legyen, ki a felelős, ha egy AI-rendszer hibás döntéseket hoz vagy nem megfelelően működik.[33]

Az elszámoltathatóság azt jelenti, hogy az AI-rendszereket tervező, fejlesztő, üzemeltető és felhasználó felek kötelesek felelősséget vállalni a rendszerek döntéseiért és azok következményeiért. Ez különösen kritikus olyan esetekben, amikor a mesterséges intelligencia rendszerek autonóm módon, minimális emberi beavatkozás mellett hoznak döntéseket, hiszen az ilyen döntések esetleges hibái vagy káros következményei széles körben éreztethetik hatásukat. Az AI rendszerek felelőssége több szinten is értelmezhető, ideértve a fejlesztési folyamatot, az adatok forrását és minőségét, valamint a rendszerek működésének felügyeletét és szabályozását is.[34]

A felelősség különböző szintjei közé tartoznak az AI-rendszereket tervező és fejlesztő szakemberek, akiknek kötelességük olyan rendszereket létrehozni, amelyek etikai alapelvek mentén működnek, és minimalizálják a hibák és torzítások lehetőségét. Emellett az AI rendszerek felhasználói és üzemeltetői is felelősséggel tartoznak, különösen akkor, ha a rendszert nem megfelelően használják, vagy nem figyelnek kellően annak működésére. A szabályozó hatóságoknak szintén alapvető szerepük van az AI-rendszerek felügyeletében, hiszen feladatuk olyan jogi és etikai szabályozási keretek létrehozása, amelyek biztosítják, hogy az AI rendszerek megfeleljenek a biztonsági és megbízhatósági elvárásoknak.[35]

A felelősségre vonásnak két alapvető típusa van: a proaktív elszámoltathatóság, amely előre meghatározott szabályok alapján történik, és a reaktív elszámoltathatóság, amely akkor lép életbe, ha már egy hiba vagy probléma történt, és visszamenőleg kell meghatározni a felelősöket. Ahhoz, hogy az AI rendszerek hosszú távon is megbízhatóak és etikailag felelősek maradjanak, világos és átfogó szabályozási keretekre van szükség, amelyek minden résztvevőt, a fejlesztőktől kezdve a felhasználókig, egyértelműen felelősségre vonhatnak a rendszerek működéséért és hatásaiért.[36]

Ez a fajta elszámoltathatóság azért is különösen fontos, mert az AI gyakran szociotechnikai rendszerek része, ahol az emberi tényezők és a technológia kölcsönhatása összetett felelősségi kérdéseket vet fel. A mesterséges intelligencia szociotechnikai szempontból történő vizsgálata a technológiai és társadalmi tényezők összefonódására összpontosít, hangsúlyozva a gépek, emberek és szervezeti környezetek közötti kölcsönhatás fontosságát. A szociotechnikai rendszerelmélet szerint az AI rendszerek bevezetése során nem csak a technikai megoldások (például gépi tanulási modellek vagy hardverek) hatékonyságára kell figyelni, hanem a társadalmi összetevőkre is, amelyek magukban foglalják a szervezeti struktúrákat, emberi interakciókat és pszichológiai tényezőket. A szociotechnikai megközelítés hangsúlyozza a technikai és társadalmi komponensek közötti harmónia szükségességét, és rámutat arra, hogy a mesterséges intelligencia rendszerek hatékonyságának

[33] Claudio, Novelli – Mariarosaria, Taddeo – Luciano, Floridi: Accountability in artificial intelligence: what it is and how it works. In AI & Soc., No. 39, 2024, 1871-1882. o.
[34] Novelli – Mariarosaria – Floridi: i.m. 1877 o.
[35] Novelli – Mariarosaria – Floridi: i.m. 1879. o.
[36] Novelli – Mariarosaria – Floridi: i.m. 1881. o.

és biztonságosságának növelése érdekében elengedhetetlen a kölcsönhatások optimalizálása. A szervezetek számára így biztosítható az AI-rendszerek megfelelő integrációja és a technikai eredmények mellett a humanisztikus célok elérése, például a kulturális és pszichológiai tényezők figyelembevételével.[37] A felelősség átláthatósága érdekében az AI rendszereket úgy kell kialakítani, hogy képesek legyenek igazolni döntéseiket és biztosítani a folyamatok ellenőrizhetőségét.

Az MI-rendszerek működésének ellenőrizhetősége és a felelősség vállalása az esetleges hibákért vagy károkért kulcsfontosságú a rendszer legitimitása szempontjából. Az MI-alapú döntésekért való felelősség meghatározása komplex feladat, amely magában foglalja a fejlesztők, az üzemeltetők és a felhasználók együttes felelősségét. A felelősségi körök világos meghatározása és a szabályozó rendszerek megfelelő alkalmazása lehetőséget biztosít a jogorvoslatra az olyan egyének vagy csoportok számára, akiket az MI döntései negatívan érintenek. Az emberi felügyelet biztosítása továbbá kulcsfontosságú a kritikus döntésekben, például az egészségügyi ellátás területén, ahol a szabályozási elvek alkalmazása és a folyamatos ellenőrzés révén az MI rendszerek hatékonysága és etikai megfelelősége is növelhető.

Igazságosság és esélyegyenlőség biztosítása: Az MI alkalmazása során kiemelt cél az igazságosság előmozdítása és az egyenlő hozzáférés biztosítása. Az MI-nek a társadalmi egyenlőség elősegítésére kell törekednie, és nem szabad növelnie a társadalmi egyenlőtlenségeket. Az ilyen rendszerek használata során külön figyelmet kell fordítani arra, hogy azok ne hozhassanak igazságtalan döntéseket, például a foglalkoztatásban, az oktatásban vagy az egészségügyi szolgáltatásokhoz való hozzáférés terén. Az AI-alapú humánerőforrás-rendszerekben például a tisztességes és átlátható működés érdekében prioritást élveznek az elfogultság minimalizálását célzó intézkedések és a GDPR szabályozásainak betartása.[38]

Kollektív felelősség és együttműködés: Az AI rendszerek fejlesztésében részt vevő minden szereplő – a fejlesztők, szabályozók, szolgáltatók és felhasználók – kollektív felelősséggel tartozik a rendszerek tisztességes és átlátható működésének biztosításáért. E közös felelősség vállalása elősegíti a technológiai átláthatóságot és a rendszerek folyamatos felügyeletét. Az együttes fellépés, amelyben az érdekelt felek együttműködnek az MI fejlesztésében és alkalmazásában, lehetővé teszi a feddhetetlenség megőrzését és a károk minimalizálását. Az ilyen modellek egyben azt is szolgálják, hogy a technológiai fejlesztések folyamatosan megfeleljenek a változó társadalmi és etikai elvárásoknak.

A mesterséges intelligencia felelős fejlesztésére és alkalmazására vonatkozó etikai elvek kidolgozásával a kutatók, kormányzati szervek és vállalatok igyekeznek kezelni az AI technológia potenciális káros hatásaival kapcsolatos növekvő aggodalmakat. Hasonlóképpen, olyan cégek, mint a Google és a Microsoft, AI-elveket fogalmaztak meg, amelyek a társadalmi előnyökre, elfogultság elkerülésére, biztonságra, átláthatóságra és elszámoltathatóságra fókuszálnak.

A *Google*, az AI felelősségteljes fejlesztésére vonatkozó alapelveinek célja, hogy biztosítsák a mesterséges intelligencia alkalmazásainak társadalmilag hasznos, etikus és biztonságos használatát. Ezen alapelvek közé tartozik, hogy az AI rendszereknek nemcsak a különböző iparágakban, például az egészségügyben, közlekedésben, energiatermelésben és szórakoztatóiparban kell pozitív hatást gyakorolniuk, hanem úgy kell működniük, hogy a társadalmi és gazdasági előnyök jelentősen meghaladják az esetleges kockázatokat. Emellett

[37] Pouria, Akbarighatar – Ilias, Pappas – Polyxeni Vassilakopoulou: A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. International Journal of Information Management Data Insights. Volume 3, Issue 2, November 2023, 100193 Elsevier, 2023, 2-3. o.

[38] Jennifer, Landrum: Fostering Ethical HR AI: Five Key Principles for Fairness, Transparency, and Compliance. (2023) https://www.linkedin.com/pulse/fostering-ethical-hr-ai-five-key-principles-fairness-landrum-ed-d (2024.07.17.)

az AI algoritmusok és adatbázisok fejlesztésekor különös figyelmet kell fordítani arra, hogy ne teremtsenek vagy erősítsenek meg elfogultságokat, különösen olyan érzékeny területeken, mint a faj, nem, vallás vagy politikai nézetek. Az AI rendszereket továbbá szigorúan kell tesztelni és biztonságos körülmények között üzembe helyezni, hogy elkerüljék a váratlan és káros következményeket, és működésüket folyamatosan felügyelni kell. Az AI rendszerekkel szemben az is elvárás, hogy biztosítsanak megfelelő elszámoltathatóságot, vagyis az emberek számára hozzáférhetőek legyenek a rendszer döntéseinek magyarázatai, lehetőséget biztosítva a visszajelzésre és fellebbezésre, ugyanakkor az AI-rendszereket megfelelő emberi irányítás alatt kell tartani. Az adatvédelem szintén kiemelt szerepet kap, amely magában foglalja, hogy a fejlesztés során tiszteletben kell tartani a felhasználók beleegyezési jogait, biztosítani kell az átláthatóságot és az adatok feletti ellenőrzést.[39]

A mesterséges intelligencia fejlesztésének magas szintű tudományos kiválóságra kell törekednie, elősegítve a tudományos kutatás és ismeretszerzés fejlődését. Ezen alapelvek betartása mellett az AI rendszerek kizárólag olyan célokra használhatók, amelyek összhangban állnak az etikai normákkal, és elkerülik a káros vagy visszaélésszerű alkalmazásokat. Az AI technológia fejlesztését elkerülik olyan területeken, ahol a technológia összességében kárt okozhat, különösen fegyverek vagy más olyan rendszerek esetében, amelyek emberek sérülését segítik elő, illetve olyan megfigyelési technológiáknál, amelyek sértik a nemzetközi normákat és az emberi jogokat.

A *Microsoft* AI-elvei a mesterséges intelligencia rendszerek felelős fejlesztését és alkalmazását szabályozzák, hangsúlyt fektetve az etikus és biztonságos működésre. Az elszámoltathatóság alapelve szerint az AI rendszereket már a fejlesztési szakaszban folyamatosan értékelni kell, különös tekintettel azok társadalmi, szervezeti és egyéni hatásaira, hogy a potenciális negatív következményeket minimalizálják. Az átláthatóság elve biztosítja, hogy a felhasználók számára egyértelművé váljon az AI rendszerek működése, különösen a döntéshozatali folyamatokban való részvételük esetén, továbbá fontos szerepet játszik a felhasználókkal való kommunikációban az AI alkalmazásának hatásairól.[40]

Az egyenlőség szempontjából a Microsoft arra törekszik, hogy az AI rendszerek fejlesztése során figyelembe vegye a különböző társadalmi csoportokat, beleértve a hátrányos helyzetűeket is, elkerülve ezzel az igazságtalan megkülönböztetést, és biztosítva, hogy mindenki egyenlő hozzáférést és minőségű szolgáltatást kapjon. A megbízhatóság és biztonság elve alapján az AI rendszereket rendszeres tesztelésnek kell alávetni, hogy biztosítsák azok megfelelő működését és a hibák minimalizálását, különösen a különböző alkalmazási környezetekben. Az adatvédelem és biztonság kulcsfontosságú az AI rendszerek fejlesztése során, hiszen ez biztosítja a személyes adatok védelmét, valamint a rendszerek integritásának fenntartását. Végezetül, a befogadás elvének célja, hogy az AI rendszerek mindenki számára hozzáférhetők legyenek, és megfeleljenek a hozzáférhetőségi normáknak, ezáltal támogatva az inkluzív technológiai fejlődést.[41]

Ezen igazságossági elvek betartása elengedhetetlen az MI felelős és fenntartható alkalmazása érdekében, mivel csak így biztosítható, hogy az MI technológia előnyei mindenki számára elérhetővé váljanak, miközben minimalizálják a potenciális kockázatokat és negatív hatásokat. A globális MI irányelvek követése és az igazságosság elveinek figyelembevétele elősegíti egy olyan mesterséges intelligencia környezet kialakítását, amely megfelel a társadalmi és etikai normáknak, és hozzájárul a társadalom bizalmának építéséhez.

[39] Google AI Principles: Objectives for building beneficial AI (2023). https://ai.google/responsibility/principles/ (2024. október 22.)
[40] Microsoft Responsible AI Standard, v2 GENERAL REQUIREMENTS (2022). https://www.microsoft.com/hu-hu/ai/principles-and-approach (2024. október 22.)
[41] Uo.

II. Magánélet és adatvédelem

Az adatvédelem és a magánélet védelme alapvető fontosságú az AI technológiák alkalmazása során, különösen az adatintenzív rendszerek esetében. Az AI által feldolgozott adatok gyakran érzékenyek, és nagy volumenben tartalmaznak személyes információkat. Az adatvédelem és a magánélet védelme szorosan összefonódik a mesterséges intelligencia (AI) fejlesztésével, különös tekintettel a nagy mennyiségű érzékeny adatok kezelésére. Az olyan technológiák, mint a differenciális adatvédelem (*differential privacy*), a federált tanulás (*federated learning*) és a homomorf titkosítás (*homomorphic encryption*) egyre fontosabbak az AI rendszerek adatvédelmi megfelelőségének biztosításában. A különböző adatvédelmi technikák, mint a federált tanulás, lehetővé teszik az AI modellek tréningjét anélkül, hogy az adatokat központilag kellene tárolni vagy megosztani, csökkentve ezzel a személyes adatok kiszivárgásának kockázatát. Ez különösen fontos az olyan érzékeny területeken, mint az egészségügy, ahol a betegek adatai védelmet igényelnek.[42]

Az informált hozzájárulás megszerzése az AI rendszerekben különösen kihívást jelent, mivel a bonyolult algoritmusok gyakran átláthatatlan adatfeldolgozást végeznek, amely megnehezíti a felhasználók számára, hogy teljes mértékben megértsék, hogyan használják fel adataikat. A dinamikus és átlátható hozzájárulási mechanizmusok segíthetnek abban, hogy a felhasználók jobban kontrollálják adataik felhasználását.[43] Ezek a technikák nemcsak az adatvédelem erősítését szolgálják, hanem elősegítik a szabályozásoknak, például az Európai Unió által bevezetett Általános Adatvédelmi Rendeletnek (GDPR) való megfelelést is.

Az egyéneknek joguk van ahhoz, hogy megvédjék személyes adataikat és magánéletüket, ezért fontos, hogy az MI alkalmazása során biztosított legyen az adatok biztonsága és a magánélet védelme. Az adatvédelmi elvek között az alábbiak szerepelhetnek:

Adat minimális felhasználása: Az MI alkalmazása során csak azokat az adatokat kell felhasználni, amelyek szükségesek a cél eléréséhez, és csak azokat az adatokat kell gyűjteni, amelyek relevánsak az adott feladathoz. Ez segít minimalizálni az adatok felhasználásából és tárolásából adódó kockázatokat. Ez az elv összhangban van a GDPR 5. cikkével, amely az adatkezelés korlátozását és az adattakarékosságot követeli meg.[44]

Átláthatóság: Az embereknek tudniuk kell, hogy milyen adatokat gyűjtenek róluk és hogyan használják fel azokat az MI rendszerek. Az átláthatóság növeli az emberek bizalmát az MI alkalmazása iránt, lehetővé teszi számukra, hogy ellenőrizzék és felügyeljék a saját adataikat. A GDPR 12. cikke előírja, hogy az adatkezelőknek átlátható módon kell tájékoztatniuk az érintetteket az adatkezelésről.[45]

Beavatkozás és kontroll: Az embereknek joguk van ahhoz, hogy befolyásolják és irányítsák a saját adataikat, valamint hogy hozzáférjenek és módosítsák azokat szükség esetén. Az MI alkalmazása során biztosítani kell az egyéneknek a lehetőséget arra, hogy kontrollálják a saját adataikat és döntéseiket. A GDPR 15-20. cikkei garantálják az érintettek jogait, beleértve az adathozzáféréshez való jogot és az adatok helyesbítéséhez való jogot.[46]

Biztonság: Az adatok biztonságát és védelmét kiemelten kell kezelni az MI alkalmazása során, hogy megelőzzük az illetéktelen hozzáférést és a személyes adatok jogosulatlan felhasználását. A GDPR 32. cikke előírja, hogy az adatkezelők és adatfeldolgozók megfelelő technikai és szervezési intézkedéseket tegyenek az adatok

[42] Nguyen, Truong – Kai, Sun – Siyao, Wang – Florian, Guitton – YiKe, Guo a: Privacy preservation in federated learning: An insightful survey from the GDPR perspective. In Computers & Security, 2021, 110. November.
[43] Nguyen – Kai – Siyao – Guitton – YiKe: i.m. 1-2. o.
[44] GDPR Általános Adatvédelmi Rendelet 5. cikk. https://gdprinfo.eu/hu/hu-article-5 (2024.07.18.)
[45] GDPR Általános Adatvédelmi Rendelet 12. cikk. https://gdprinfo.eu/hu/hu-article-12 (2024.07.18.)
[46] GDPR Általános Adatvédelmi Rendelet 15-20. cikkei. https://gdprinfo.eu/hu (2024.07.18.)

biztonsága érdekében. [47]

Az adatvédelmi elvek betartása és az emberek jogainak tiszteletben tartása elengedhetetlen az MI felelős és fenntartható alkalmazásában, és segít abban, hogy az MI előnyei maximálisan kihasználhatók legyenek, miközben minimalizálják a kockázatokat és negatív hatásokat.

II.1. Adatgyűjtés- és feldolgozás korlátai

Az adatgyűjtés- és feldolgozás során számos korláttal kell szembenéznünk, különösen a mesterséges intelligencia alkalmazása során, mint például az alábbiak:

Adatok minősége és hozzáférhetősége: Nem minden adat áll rendelkezésre az MI számára, az elérhető adatok minősége és mennyisége változó lehet. Az adatok gyakran hiányosak, torzítottak vagy nem reprezentatívak, ami jelentősen csökkentheti az MI hatékonyságát és megbízhatóságát.[48] A nem megfelelő adatminőség pontatlan vagy elfogult eredményekhez vezethet, ami befolyásolhatja az MI alkalmazásainak döntéseit. Például, egyes adathalmazok hiányosak lehetnek vagy nem tartalmazhatják az összes releváns információt, ami korlátozza az MI modellek képességeit.[49] Az alacsony adatminőség számos okra vezethető vissza, például emberi hibákra, érvénytelen információkra, hibás gépi feldolgozásra, strukturálatlan adatokra és hiányzó értékekre. Ezek a problémák különösen fontosak a Big Data alkalmazások esetében, ahol a hatalmas adatmennyiség (Volume) kezelése új, komplex kihívásokat teremt a rendszerek számára, például a méretezhetőség és a sebesség tekintetében. A hibás adatok komoly negatív hatással lehetnek a döntéshozatali folyamatokra, növelve a működési költségeket, és akadályozva a szervezeti stratégiák hatékony megvalósítását, ami végső soron kihat a versenyképességre is. A megfelelő adatminőség-menedzsment elengedhetetlen ahhoz, hogy a nagy adatállományok kezelése hatékonyan történjen, különös figyelmet fordítva az adatok biztonsági mentésére, helyreállítására, valamint a hibás adatok szűrésére. Az adatminőség közvetlen hatással van a szervezetek működésére, hiszen a pontatlan adatok nemcsak az ügyfélbizalmat és elégedettséget csökkenthetik, hanem rontják a termelékenységet és növelik a pénzügyi kockázatokat is.[50]

Adatvédelmi szabályozások: Az adatvédelmi törvények és szabályozások, mint például az Általános Adatvédelmi Rendelet (GDPR)[51], korlátozhatják az adatgyűjtést- és feldolgozást, különösen, ha személyes adatokról van szó. Ezek a szabályozások szigorú követelményeket írnak elő az adatok gyűjtésére, tárolására és feldolgozására vonatkozóan, és befolyásolhatják az MI képességét a releváns adatok gyűjtésére és használatára. Az adatvédelmi törvények betartása elengedhetetlen az egyének magánéletének védelme érdekében, de egyben korlátozhatja az MI alkalmazások fejlesztésének és működtetésének lehetőségeit.[52]

Adatok torzítása és elfogultsága: Az adatok gyakran torzítottak lehetnek előítéletek vagy diszkrimináció miatt, ami befolyásolhatja az MI által hozott döntéseket. Például, ha az adatok csak egy bizonyos csoportot képviselnek, az MI által hozott döntések elfogultak lehetnek ezen csoportok másokkal szemben. Az adatokban rejlő elfogultságok és torzítások továbbörökíthetik a társadalmi egyenlőtlenségeket, és igazságtalan eredményekhez

[47] GDPR Általános Adatvédelmi Rendelet 32. cikk. https://gdprinfo.eu/hu/hu-article-32 (2024.07.18.)
[48] Abdallah, Mohammad – Muhairat, Mohammad – Althunibat, Ahmad – Abdalla, Ayman: Data Quality: Factors, Frameworks, and Challenges. In COMPUSOFT, No. 8, 2020, 3787. o.
[49] Carina, Prunkl – Jess, Whittlestone: Beyond Near- and Long-Term: Towards a Clearer Account of Research Priorities in AI Ethics and Society. https://doi.org/10.48550/arXiv.2001.04335 (2024.07.18.)
[50] Abdallah – Muhairat – Althunibat – Abdalla: i.m. 3787-3788. o.
[51] Lásd bővebben: General Data Protection Regulation (GDPR): https://gdpr.eu/ (2024.07.18.)
[52] Lásd bővebben: GDPR Information Portal: https://www.privacy-regulation.eu/en/index.htm (2024.07.18.)

vezethetnek az MI alkalmazásain keresztül.[53]

Szabványok és kompatibilitás: Az adatok gyakran különböző formátumokban és forrásokból származnak, ami nehezítheti azok összegyűjtését és összehasonlítását. Az adatok összeillesztése és integrálása kihívást jelenthet az MI rendszerek számára. Az adatok interoperabilitásának[54] hiánya akadályozhatja az MI rendszerek hatékony működését, és korlátozhatja a modellek képességét az adatokból történő pontos következtetések levonására.

Ezek a korlátok felhívják a figyelmet arra, hogy az adatgyűjtés- és feldolgozás nem mindig egyenletes és egyszerű folyamat, és számos tényezőt kell figyelembe venni az adatok használata során. Az MI fejlesztőinek és felhasználóinak tisztában kell lenniük ezekkel a korlátokkal, és megfelelő intézkedéseket kell tenniük az adatok minőségének és megbízhatóságának biztosítása érdekében.

II.3. Felhasználói adatok védelmének módszerei

A felhasználói adatok védelmének számos módszere között kulcsfontosságú szerepet játszanak az adatvédelmi szabályzatok és jogszabályok betartása, melyek megfelelnek az adatvédelmi törvényeknek és szabályozásoknak. Ezek az irányelvek meghatározzák az adatgyűjtés- és feldolgozás legmegfelelőbb gyakorlatait, valamint az adatok tárolásának, kezelésének és hozzáférésének szabályait.[55] Az adatvédelmi elvek alkalmazása különösen fontos az adatvédelmi törvények és szabályozások, például az Európai Unió Általános Adatvédelmi Rendelete (GDPR)[56] és az Egyesült Államokban a CCPA (California Consumer Privacy Act)[57] által előírt követelmények teljesítése érdekében. Emellett az adatvédelem terén alapvető elv az adatminimalizálás, melynek lényege, hogy csak azokat az adatokat gyűjtsük és tároljuk, amelyek elengedhetetlenek az adott feladat végrehajtásához. Ez segít minimalizálni az adatokhoz kapcsolódó kockázatokat, és csökkenti az esetleges biztonsági sérülések vagy adatvesztés veszélyét. A GDPR 5. cikke pontja kimondja, hogy a személyes adatoknak az adatkezelés céljai szempontjából megfelelőnek és relevánsnak kell lenniük, a szükséges mértékre kell korlátozódniuk.[58]

Az adatvédelmi intézkedések során kiemelt figyelmet kell fordítani az erősített biztonsági intézkedések bevezetésére, melyek magukban foglalhatják a titkosítást, a tűzfalakat, a többrétegű hitelesítést és az adatvédelmi szoftverek használatát. Ezek a technikák és eszközök segítenek megvédeni az adatokat a jogosulatlan hozzáféréstől és a biztonsági fenyegetésektől, biztosítva ezzel az adatok biztonságát és védelmét. Az adatvédelem másik fontos eleme a felhasználói beleegyezés kérése az adatok gyűjtéséhez és felhasználásához. A felhasználók tudatában kell lenniük annak, hogy milyen adatokat gyűjtenek róluk, és milyen célból használják fel azokat az MI rendszerek. A GDPR 7. cikke előírja, hogy a

---

[53] Barocas, Solon – Hardt, Moritz – Narayanan, Arvind: Fairness and machine learning. Limitations and Opportunities (2023). 10. o. https://fairmlbook.org/pdf/fairmlbook.pdf (2024.07.18.)

[54] Interoperabilitás: Együttműködési képességet jelent, mely az Európai Unió Interoperabilitási Keretrendszere szerint egyaránt értelmezhető politikai, jogi szervezeti, szemantikai és technikai értelemben. Lásd bővebben: https://lexikon.uni-nke.hu/ (2024.07.18.)

[55] Kuner, Christopher – Bygrave, Lee, A – Docksey, Christopher – Drechsler, Laura: The EU General Data Protection Regulation (GDPR): A Commentary. Oxford University Press, 2020. 100-120. o.

[56] GDPR Information Portal: https://www.privacy-regulation.eu/en/index.htm (2024.07.18.)

[57] Lásd bővebben: California Consumer Privacy Act (CCPA): https://oag.ca.gov/privacy/ccpa (2024.07.18.)

[58] Voigt, Paul – Bussche, Axel von dem: The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer International Publishing, 2017, 187. o., Paul Voigt és Axel von dem Bussche *The EU General Data Protection Regulation (GDPR): A Practical Guide* című könyvében részletesen tárgyalja az adatminimalizálás elvét, amely központi szerepet játszik az adatvédelemben. A szerzők hangsúlyozzák, hogy a GDPR 5. cikkének (1) bekezdése c) pontja szerint a személyes adatoknak „megfelelőnek, relevánsnak és a szükséges mértékre korlátozottnak" kell lenniük az adatkezelés céljainak eléréséhez.

beleegyezésnek egyértelműnek és önkéntesnek kell lennie, valamint biztosítani kell a felhasználók számára az adataikhoz való hozzáférés, azok módosítása és törlése jogát. Az átláthatóság növeli a felhasználói bizalmat és lehetővé teszi számukra, hogy ellenőrizzék és irányítsák saját adataikat.

A rendszeres ellenőrzések és frissítések szintén elengedhetetlenek az adatvédelmi intézkedések hatékonyságának biztosítása érdekében. Ezek az ellenőrzések lehetővé teszik a biztonsági rések azonosítását és javítását, valamint az adatvédelmi rendszerek és eljárások frissítését az új fenyegetések és kihívások kezelése érdekében. Az ISO/IEC 27001 szabvány[59] irányelvei alapján az információbiztonsági irányítási rendszerek rendszeres auditálása és karbantartása elengedhetetlen. A felhasználók oktatása és tájékoztatása szintén fontos része az adatvédelmi stratégiának, mivel ezek az emberek lehetnek az első vonalban a védelem és az adatvédelem kérdéseinek kezelése során. A National Institute of Standards and Technology (NIST)[60] irányelvei hangsúlyozzák a felhasználói tudatosság növelésének fontosságát az információbiztonság javítása érdekében.

Ezek az adatvédelmi módszerek és gyakorlatok alapvetőek a felhasználói adatok biztonságának és védelmének biztosítása érdekében. A hatékony adatvédelem elősegíti a bizalom kialakulását a felhasználók körében, miközben megfelel a jogszabályi előírásoknak és minimalizálja a biztonsági kockázatokat.

III. Emberi méltóság és személyiségi jogok

Az emberi méltóság és a személyiségi jogok központi szerepet töltenek be a mesterséges intelligencia alkalmazása során, és különös figyelmet igényelnek az adatgyűjtés, adatfeldolgozás és döntéshozatal minden szakaszában. Az emberi méltóság elve kimondja, hogy minden egyén joga, hogy tiszteletben tartsák méltóságát és emberi jogait, és ezeket az alapelveket be kell építeni minden MI technológiába.[61]

Az emberi méltóság és a személyiségi jogok védelme azt jelenti, hogy az embereknek joguk van a magánélethez és személyes adataik védelméhez. Ez magában foglalja a személyes adatok gyűjtésének és felhasználásának korlátozását, valamint adatvédelmi intézkedések és eljárások bevezetését az adatok biztonságának és magánéletük védelmének biztosítása érdekében. Az Emberi Jogok Egyetemes Nyilatkozata (ENSZ, 1948)[62] és az Európai Emberi Jogi Egyezmény (EJEE, 1950)[63] szintén alapvető elveket fektetnek le az emberi méltóság és a személyiségi jogok védelmére.

A méltóság és személyiségi jogok tiszteletben tartása azt is jelenti, hogy az MI által meghozott döntéseket etikusan és felelősen kell végrehajtani, figyelembe véve az emberek érdekeit és jogaikat. Ezeknek a döntéseknek összhangban kell lenniük az emberi jogokkal és

---

[59] Lásd bővebben: ISO 27001 Információbiztonsági Irányítási Rendszer (IBIR). https://www.tuvsud.com/hu-hu/szolgaltatasok/audit-es-rendszertanusitas/iso-27001 (2024.07.18.)

[60] National Institute of Standards and Technology (NIST) Cybersecurity Framework honlapja: https://www.nist.gov/cyberframework (2024.07.18.)

[61] Alaptörvény II. Cikk: https://net.jogtar.hu/jogszabaly?docid=a1100425.atv (2024. október 18.)

[62] Az ENSZ Közgyűlése 1948-ban fogadta el, és az emberi jogok egyetemességét hangsúlyozza. Az emberi méltóság és az alapvető jogok védelmét a Nyilatkozat számos cikkelye tartalmazza, például az 1. és a 3. cikkben, amelyek kifejezetten az emberi méltóságot és az élethez, szabadsághoz való jogot védik. Emberi Jogok Egyetemes Nyilatkozata (teljes szöveg): https://www.coe.int/hu/web/compass/the-universal-declaration-of-human-rights-full-version- (2024.07.18.)

[63] Az Európa Tanács által 1950-ben elfogadott Egyezmény az emberi jogok és alapvető szabadságok védelméről szól, és jogi keretet biztosít az európai polgárok személyiségi jogainak védelméhez. Az 1. cikk garantálja az Egyezményben foglalt jogok és szabadságok védelmét, míg a 8. cikk különösen a magán- és családi élet tiszteletben tartásához való jogot hangsúlyozza. Az emberi jogok európai egyezményének honlapja: https://www.coe.int/en/web/human-rights-convention (2024.07.18.)

az etikai elvekkel, és nem szabad sérteniük vagy hátrányosan befolyásolniuk az egyének méltóságát és jogait.

Mindezek elengedhetetlenek a felelős és fenntartható mesterséges intelligencia alkalmazásához, továbbá segítenek abban, hogy a technológia előnyeit maximálisan kihasználhassuk, miközben minimalizáljuk a kockázatokat és a negatív hatásokat az emberekre és a társadalomra.

III.1. Az MI emberi méltóságára gyakorolt hatásai

Az MI emberi méltóságra gyakorolt hatásai sokrétűek és összetettek, a technológia különböző alkalmazási módjaitól függően változhatnak. Az autonóm döntéshozatal például azzal járhat, hogy az emberek elveszítik az irányítást és a kontrollt saját életük és döntéseik felett. Ez különösen kritikus lehet az egészségügyben vagy a jogi rendszerekben, ahol az MI által hozott döntések jelentős hatással lehetnek az egyének életére.

Az algoritmusok által vezérelt diszkrimináció és előítéletesség is komoly probléma lehet. Ha az MI döntései torzított vagy előítéletes adatokon alapulnak, az emberek diszkriminációnak vagy igazságtalanságnak lehetnek kitéve, különösen a munkahelyi döntéshozatalban vagy a bűnüldözési rendszerekben. Alkalmazása a személyes kapcsolatok hiányához és társadalmi elszigeteltséghez is vezethet. Az automatizált rendszerek, mint a chatbotok vagy az ügyfélszolgálati automaták, egyre inkább átveszik az emberi interakció szerepét, ami csökkentheti az emberi érintettség és empátia szintjét, és növelheti a társadalmi elszigeteltséget. Az MI által gyűjtött és feldolgozott adatok nagy mennyisége és pontossága veszélyeztetheti az egyének személyazonosságát és magánéletét. A személyes adatok védelme és a magánélet megőrzése ezért kritikus fontosságú, hogy biztosítsuk az egyének bizalmát és méltóságát.

Az *AI és a diszkrimináció* kérdése komoly kihívásokat vet fel, mivel az emberi előítéletek gyakran átszivárognak a mesterséges intelligencia rendszerekbe. Ezek a rendszerek nem semlegesek, hiszen az emberek által betanított adatokra építenek, amelyek maguk is torzultak lehetnek. Az MI rendszerek esetében többféle torzítással kell szembenézni, például implicit (tudattalan) előítélettel, minták torzításával (sampling bias), időbeli torzítással (temporal bias), és a tréning adatok túlzott illesztésével. Mindezek komoly hatással lehetnek a mindennapi élet számos területén, beleértve az egészségügyi és munkaügyi döntéseket is. A megoldások között fontos szerepe van az inkluzív tervezésnek és a felhasználói tesztelésnek, amelyekkel csökkenthető a diszkrimináció kockázata. Ezen kívül, az algoritmusokat különböző környezetekben kell tesztelni, hogy kiderüljön, általánosíthatók-e szélesebb körben. Az igazságosabb MI rendszerek fejlesztése érdekében szorosabb ellenőrzést és mélyebb adatvizsgálatot igényelnek ezek az eszközök.[64]

A mesterséges intelligencia és gépi tanulás rendszerek gyors fejlődése számos adatvédelmi és etikai kihívást vet fel. A magánélet védelmével foglalkozó szakembereknek érteniük kell a technológiai rendszerek működését ahhoz, hogy megfelelően alkalmazzák a vonatkozó adatvédelmi jogszabályokat, például a GDPR-t. Az AI rendszereknél az átláthatóság, az igazságosság, a biztonság és az elszámoltathatóság kulcsfontosságú szempontok, amelyeket be kell tartani. Fontosak a globális szinten kialakított szabályozási keretek, mint például az UNESCO és az OECD AI-elveit, valamint a gyakorlatban alkalmazott vállalati irányelveket. Számos ország, mint az Egyesült Államok és az Európai Unió, már elkezdett dolgozni saját mesterséges intelligencia szabályozási keretein, és az AI felelős használatának biztosítása érdekében fokozott figyelmet szentelnek a jogszabályoknak

[64] Emerging Technologies (2021): Research shows AI is often biased. Here's how to make algorithms work for all of us https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/ (2024. október 29.)

és etikai kérdéseknek.[65]

Ezek a hatások kiemelik az emberi méltóság és az egyéni jogok védelmének fontosságát az MI rendszerek alkalmazása során. Az etikai irányelvek és szabályozások betartása, valamint a társadalmi párbeszéd és az átláthatóság elősegítése fontosak az MI emberi méltóságára gyakorolt hatásainak kezelésében és minimalizálásában.

III.2. Személyes adatok és jogi védelem

A mesterséges intelligencia technológiák alkalmazása során a személyes adatok védelme és a jogi védelem kritikus jelentőségű, különösen, ha nagy mennyiségű és érzékeny adatot használnak fel. A GDPR hatálya alá tartozó területeken az adatkezeléshez kapcsolódó szabályozások szigorúan meghatározzák az adatgyűjtés- és feldolgozás feltételeit, biztosítva az egyének adatainak védelmét. Az adatminimalizálás elve, mely szerint csak a szükséges adatokat gyűjtik és tárolják, szintén kiemelt jelentőségű az adatbiztonság fenntartásában.[66]

Az adatbiztonság érdekében kiemelten fontos fejlett technológiai védelmi eszközök bevezetése, amelyek magukban foglalják az adatok titkosítását, a tűzfalak telepítését, a többlépcsős hitelesítési folyamatok alkalmazását, valamint a speciális adatvédelmi szoftverek integrálását. Ezen biztonsági megoldások alapvető funkciója, hogy meggátolják az illetéktelen hozzáférést és minimalizálják a kibertámadások vagy adatszivárgások kockázatát, ezáltal fenntartva az adatok integritását és bizalmasságát. Az egyének számára elengedhetetlen, hogy átlátható tájékoztatást kapjanak arról, milyen személyes adatokat gyűjtenek róluk, és ezen adatok milyen célra kerülnek felhasználásra. Az adatkezelés átláthatósága és a hozzáférési jogok biztosítása szorosan összefügg az egyének adatvédelmi jogainak védelmével, különös tekintettel az adatok módosításának és törlésének lehetőségére. Az átláthatóság és az elszámoltathatóság az adatvédelmi irányelvek központi eleme, amely növeli a felhasználói bizalmat és támogatja az etikus adatkezelést. A biztonsági rendszerek folyamatos fejlesztése és karbantartása elengedhetetlen annak érdekében, hogy reagálni lehessen az újonnan felmerülő fenyegetésekre. Ehhez rendszeres auditok és biztonsági felülvizsgálatok szükségesek, amelyek célja az adatvédelmi intézkedések naprakészségének és hatékonyságának fenntartása a változó technológiai környezetben.

Az olyan technikai megoldások, mint a titkosítás, a tűzfalak és a többrétegű hitelesítés, kulcsszerepet játszanak az adatvédelemben. A jogi megfelelés érdekében az egyének jogait, például az adatokhoz való hozzáférést, módosítást és törlést, szigorúan tiszteletben kell tartani. Az adatvédelmi szakemberek feladata ezeknél a rendszereknél az adatvédelmi kockázatok elemzése és a jogszabályok betartásának biztosítása. A rendszer tervezési céljainak, az adatok minőségének, valamint a fejlesztési és tesztelési folyamatok alapos átgondolása. Az átláthatóság és elszámoltathatóság is kulcsfontosságú a kockázatminimalizálásban.[67] Az ilyen adatvédelmi szabályozások nemcsak az EU területén, hanem világszerte, többek között az Egyesült Államokban és Kínában is alapvető szerepet töltenek be az MI rendszerek felelősségteljes alkalmazásában.

Összességében az adatvédelem és a jogi védelem alapvető fontosságúak a mesterséges intelligencia alkalmazása során, hogy védve legyen az egyének személyes adatai és

[65] Koerner, Katharina: Privacy and responsible AI. IAPP (2022) https://iapp.org/news/a/privacy-and-responsible-ai/ (2024. október 19.)

[66] FRA Bias in Algorithms – Artificial Intelligence and Discrimination. Vienna (2022) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf (2024. október 29.)

[67] Marvin, van Bekkum – Frederik, Zuiderveen: Using sensitive data to prevent AI discrimination: Does the EU GDPR need a new exception? IAPP (2023) https://iapp.org/news/a/using-sensitive-data-to-prevent-ai-discrimination-does-the-eu-gdpr-need-a-new-exception/ (2024. október 19.)

magánélete, és biztosítva legyen az adatok biztonsága az adatgyűjtés- és feldolgozás minden szakaszában.

IV. Megállapítások

A mesterséges intelligencia etikai kihívásai között kiemelkedő szerepe van az adatvédelem és a magánélet védelmének, mivel az MI rendszerek gyakran nagy mennyiségű érzékeny adatot dolgoznak fel. Az egyének személyes adatai veszélybe kerülhetnek, ha nem biztosítják azok megfelelő védelmét, ami különösen fontos az olyan technológiák esetében, amelyek automatizált döntéshozatalt alkalmaznak. Az adatvédelmi szabályzatok, mint például az Európai Unió általános adatvédelmi rendelete (GDPR), szigorú előírásokat tartalmaznak az adatok gyűjtésére, kezelésére és tárolására vonatkozóan. Ezen belül az adatminimalizálás elve kulcsfontosságú, amely szerint csak azokat az adatokat szabad gyűjteni és feldolgozni, amelyek feltétlenül szükségesek az adott cél eléréséhez. Ez az elv segít csökkenteni az adatbiztonsági kockázatokat és a személyes adatokkal kapcsolatos visszaélések esélyét. A technikai megoldások, mint a titkosítás, tűzfalak és többrétegű hitelesítés, szintén alapvetőek az adatok biztonságának biztosításában. A személyes adatok védelme nemcsak jogi követelmény, hanem erkölcsi kötelezettség is, amely elősegíti a felhasználói bizalom fenntartását és az AI rendszerek társadalmi elfogadását. Az AI alkalmazása során nemcsak az adatvédelem, hanem az átláthatóság és az elszámoltathatóság is kulcsfontosságú szerepet játszik, hogy minimalizálni lehessen a kockázatokat és biztosítani lehessen a jogi megfelelést.

A mesterséges intelligencia adatvédelmi szempontú vizsgálata során a jövőbeli kutatások egyik kiemelt feladata egy olyan átfogó szabályozási és etikai keretrendszer megalkotása, amely figyelembe veszi az MI rendszerek dinamikus fejlődési és tanulási képességeit. A jelenlegi jogi szabályozások, mint például a GDPR, már tartalmaznak irányelveket a személyes adatok védelmére és az átláthatóság biztosítására, azonban ezek a szabályok nem mindig elegendőek a gépi tanulás összetett természetének kezelésére.

Eddigi kutatásaim alapján javaslom, hogy az adatvédelmi szabályozásokban jelenjen meg az algoritmusok *felelősségi és átláthatósági folyamatos auditálása*, amely nem csak statikus elemzésekre épít, hanem valós idejű visszacsatolási mechanizmusokat is integrál a rendszerek működésének értékelésébe. Ez nem csupán a felhasználói bizalom fenntartását segíti elő, hanem biztosítja azt is, hogy az MI alkalmazások ne mélyítsék tovább a társadalmi egyenlőtlenségeket azáltal, hogy rejtett torzításokat rögzítenek a döntéshozatal során. Az MI technológiák előrehaladott alkalmazási területein, mint a bűnüldözés, a foglalkoztatás vagy az egészségügy, a jogi szabályozásoknak figyelembe kell venniük az *adatok dinamikus jellegét*, különös tekintettel az algoritmusok adaptációs képességére és a kontextusfüggő adatok változékonyságára. Ezzel a dinamikus keretrendszerrel nemcsak a technológiai, hanem az etikai megfelelőség is garantálható, amely az egyén alapvető jogainak védelmét szolgálja egy folyamatosan változó digitális környezetben. Ez az állítás nemcsak a meglévő kutatási eredmények összegzését, hanem egy jövőbeli megoldás felé mutató innovatív gondolkodásmódot is tükröz, amely az MI technológiák hosszú távú társadalmi és jogi fenntarthatóságának kulcsa.

Az adatvédelem és a mesterséges intelligencia összefonódása nem csupán technikai feladat, hanem egy mélyreható társadalmi és erkölcsi kötelezettség, amely megkívánja, hogy a jogalkotók és kutatók új szabályozási kereteket és megoldásokat dolgozzanak ki a digitális jövő fenntarthatósága érdekében. Egy társadalom akkor lehet igazán sikeres, ha képes a technológiai haladást és az egyéni jogok védelmét harmonikusan összehangolni a folyamatos változások közepette. A mesterséges intelligencia nem válhat uralkodó eszközzé; feladata az, hogy az emberi érdekeket szolgálva egyenlőbb, biztonságosabb és emberségesebb világot

segítsen teremteni. Az MI igazi sikere abban rejlik, hogy nemcsak technológiai vívmányokkal járul hozzá a fejlődéshez, hanem társadalmi és erkölcsi értelemben is az emberiség jólétét támogatja.

Irodalomjegyzék

A rasszizmus és intolerancia elleni európai bizottság 7. Sz. Általános ajánlása: a rasszizmus és a faji megkülönböztetés elleni küzdelem a nemzeti jogalkotásban, https://rm.coe.int/ecri-general-policy-recommendation-no-7-revised-on-national-legislatio/16808b5ab3 (2024. október 16.)

Abdallah, Mohammad – Muhairat, Mohammad – Althunibat, Ahmad – Abdalla, Ayman: Data Quality: Factors, Frameworks, and Challenges, In COMPUSOFT, An international journal of advanced computer technology, No 8, 2020. https://www.researchgate.net/publication/344224569_Big_Data_Quality_Factors_Frameworks_and_Challenges (2024.07.18.)

Magyarország Alaptörvénye

Alexander, Michelle: The New Jim Crow: Mass Incarceration in the Age of Colorblindness. New Press, 2012. (2024. október 19.)

Andersen, Grady – MoldStud Research Team: Ethics in Artificial Intelligence Systems Analysis: Ensuring Fairness and Accountability. Legal Frameworks for Ensuring Accountability in Artificial Intelligence Systems. (2024). https://moldstud.com/articles/p-ethics-in-artificial-intelligence-systems-analysis-ensuring-fairness-and-accountability (2024.október 21.)

Ankarstad, Nicklas: What is Explainable AI (XAI)? In Medium, 2020. https://towardsdatascience.com/what-is-explainable-ai-xai-afc56938d513 (2024. október 16.)

Ashraf, Afsa: Why Artificial Intelligence Requires Human Intervention. (2022) https://www.royalcyber.com/blogs/ai-systems-need-human-intervention/ (2024.07.17.)

Az OECD Multinacionális vállalatokra vonatkozó irányelvei (Irányelvek): https://oecdmnkp.hu/hu/iranyelvek (2024. október 17.)

Barocas, Solon – Hardt, Moritz – Narayanan, Arvind: Fairness and machine learning. Limitations and Opportunities, 2023. https://fairmlbook.org/pdf/fairmlbook.pdf (2024.07.18.)

Bekkum, Marvin van – Zuiderveen, Frederik: Using sensitive data to prevent AI discrimination: Does the EU GDPR need a new exception? IAPP, 2023. https://iapp.org/news/a/using-sensitive-data-to-prevent-ai-discrimination-does-the-eu-gdpr-need-a-new-exception/ (2024. október 19.)

Capitol Technology University: The Ethical Considerations of Artificial Intelligence, 3. bek. (2023) https://www.captechu.edu/blog/ethical-considerations-of-artificial-intelligence (2024. október 16.)

Centre for Information Policy Leadership (CIPL) (2020): Artificial Intelligence and Data Protection How the GDPR Regulates AI. https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (2024.október 21.)

Chengeta, Thompson: Accountability gap: Autonomous weapon systems and modes of responsibility in international law. In Denver Journal of International Law and Policy, 2016/1. https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1011&context=djilp https://doi.org/10.2139/ssrn.2755211

Corbett-Davies, Sam – Pierson, Emma – Avi, Feller – Sharad, Goel – Aziz, Huq: Algorithmic decision making and the cost of fairness. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13-17 August 2017. https://dl.acm.org/doi/abs/10.1145/3097983.3098095 (2024. október 18.)

Crawford, Neta, C.: Individual and Collective Moral Responsibility for Systemic Military Atrocity. In Journal of Political Philosophy, 2007/2. 187-212. https://doi.org/10.1111/j.1467-9760.2007.00278.x

Danaher, John: Tragic Choices and the Virtue of Techno-Responsibility Gaps. In Philosophy & Technology, 2022/2. https://doi.org/10.1007/s13347-022-00519-1.

Emberi Jogok Egyetemes Nyilatkozata https://www.coe.int/hu/web/compass/the-universal-declaration-of-human-rights-full-version- (2024.07.18.)

Emerging Technologies, 2021: Research shows AI is often biased. Here's how to make algorithms work for all of us https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/ (2024. október 29.)

Ethics guidelines for trustworthy AI (2019). European Comission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2024.07.17.)

Európai Bizottság hivatalos honlapja: https://commission.europa.eu/index_hu (2024. október 17.)

European Parliament: Understanding algorithmic decision-making: Opportunities and challenges. STOA | Panel for the Future of Science and Technology. Scientific Foresight Unit (STOA) (2019). https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf (2024. július 24.)

Felzmann, Heike – Fosch-Villaronga, Eduard – Lutz, Christoph – Tamó-Larrieux, Aurelia: Towards Transparency by Design for Artificial Intelligence. In Sci Eng Ethics, No. 26, 2020. https://doi.org/10.1007/s11948-020-00276-4 (2024. október 21.)

FRA Bias in Algorithms – Artificial Intelligence and Discrimination. Vienna, 2022. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf (2024. október 29.)
GDPR Általános Adatvédelmi Rendelet https://gdprinfo.eu/hu (2024.07.18.)

GDPR Information Portal: https://www.privacy-regulation.eu/en/index.htm (2024.07.18.)

GDPR: Az európai parlament és a tanács 2016. április 27-i (EU) 2016/679 rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet)

General Data Protection Regulation (GDPR): https://gdpr.eu/ (2024.07.18.)

Glenn, Gordon: The Use of Artificial Intelligence in the Legal Profession https://www.lexisnexis.com/community/insights/legal/practical-guidance-journal/b/pa/posts/the-use-of-artificial-intelligence-in-the-legal-profession?srsltid=AfmBOorWKjTDsoYam uuR2leSQfzwQqGoKodMTdevfkX2_AwrHllspz5 (2024. október 21.)

Google AI Principles: Objectives for building beneficial AI (2023). https://ai.google/responsibility/principles/ (2024. október 22.)

Heinrichs, Bert: Discrimination in the age of artificial intelligence. In AI & Society, No. 37, 2022. https://doi.org/10.1007/s00146-021-01192-2 (2024. október 21.)

https://op.europa.eu/hu/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (2024. október 19.)

Institute of Electrical and Electronics Engineers (IEEE) hivatalos honlapja: https://www.ieee.org/ (2024. október 19.)

Institute of Electrical and Electronics Engineers Standards Association (IEEE SA): Autonomous and Intelligent Systems (AIS) Standards. https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/ (2024. október 19.)

ISO 27001 Információbiztonsági Irányítási Rendszer (IBIR). https://www.tuvsud.com/hu-hu/szolgaltatasok/audit-es-rendszertanusitas/iso-27001 (2024.07.18.)

Kleinberg, Jon – Ludwig, Jens – Mullainathan, Sendhil – Sunstein, Cass, R: Discrimination in the Age of Algorithms. In Journal of Legal Analysis, No. 10, 2018. 113-174. 116. o. https://doi.org/10.1093/jla/laz001 (2024. október 19.)

Koerner, Katharina: Privacy and responsible AI. IAPP, 2022. https://iapp.org/news/a/privacy-and-responsible-ai/ (2024. október 19.)

Kuner, Christopher – Bygrave, Lee, A – Docksey, Christopher – Drechsler, Laura: The EU General Data Protection Regulation (GDPR): A Commentary. Oxford University Press, 2020. https://doi.org/10.1093/oso/9780198826491.001.0001 (2024.07.18.)

Landrum, Jennifer: Fostering Ethical HR AI: Five Key Principles for Fairness, Transparency, and Compliance. (2023) https://www.linkedin.com/pulse/fostering-ethical-hr-ai-five-key-principles-fairness-landrum-ed-d (2024.07.17.)

California Consumer Privacy Act (CCPA): https://oag.ca.gov/privacy/ccpa (2024.07.18.)

Mehdi, Dastani – Vahid, Yazdanpanahof: Responsibility of AI Systems. In AI & Soc, 2023. 38. https://doi.org/10.1007/s00146-022-01481-4 (2024. október 17.)

Mesterséges intelligenciával foglalkozó magas szintű független szakértői csoport (2019): Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan. doi:10.2759/428483

Microsoft Responsible AI Standard, v2 GENERAL REQUIREMENTS (2022). https://www.microsoft.com/hu-hu/ai/principles-and-approach (2024. október 22.)

National Institute of Standards and Technology (NIST) Cybersecurity Framework honlapja: https://www.nist.gov/cyberframework (2024.07.18.)

Nguyen, Truong – Kai, Sun – Siyao, Wang – Florian, Guitton – YiKe, Guo a: Privacy preservation in federated learning: An insightful survey from the GDPR perspective. In Computers & Security, 2021. 110. https://doi.org/10.1016/j.cose.2021.102402 (2024. október 22.)

Novelli, Claudio – Taddeo, Mariarosaria – Floridi, Luciano: Accountability in artificial intelligence: what it is and how it works. In AI & Soc, 2024, 39. https://doi.org/10.1007/s00146-023-01635-y (2024. október 22.)

Pouria, Akbarighatar – Ilias, Pappas – Polyxeni Vassilakopoulou: A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. In International Journal of Information Management Data Insights, No. 2, 2023. https://doi.org/10.1016/j.jjimei.2023.100193 (2024. október 22.)

Prunkl, Carina – Whittlestone, Jess: Beyond Near- and Long-Term: Towards a Clearer Account of Research Priorities in AI Ethics and Society. https://doi.org/10.48550/arXiv.2001.04335 (2024.07.18.)

S.2892 – Algorithmic Accountability Act of 2023: https://www.congress.gov/bill/118th-congress/senate-bill/2892/text (2024. október 16.)

The Ethical Considerations of Artificial Intelligence, Capitol, Bias and Discrimination (2023), Capitol Technology University https://www.captechu.edu/blog/ethical-considerations-of-artificial-intelligence (2024. október 16.)

UNESCO: Recommendation on the Ethics of Artificial Intelligence: Respect, protection and promotion of human rights and fundamental freedoms and human dignity. 2021. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (2024. október 17.)

Voigt, Paul – Bussche, Axel von dem: The EU General Data Protection Regulation (GDPR): A Practical Guide. 1st Edition., Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-57959-7 (2024.07.18.)