Németh Zoltán*

A mesterséges intelligencia és az algoritmikus diszkrimináció jogi kihívásai

1. BEVEZETÉS

A mesterséges intelligencia (a továbbiakban: MI) és az algoritmusok által vezérelt automatizált döntéshozatal napjainkra az élet szinte minden területén megjelent. Az MI-alapú rendszerek képesek elősegíteni az egészségügyi diagnosztikát, hatékonyabbá tenni a bűnüldözést, biztonságosabbá tenni a közlekedést, vagy éppen felismerni a csalásokat és a kibertámadásokat.[1] Ugyanakkor ezen technológiák rohamos térnyerése komoly jogi és társadalmi kihívásokat is felvet. Különös aggodalomra ad okot az a jelenség, amikor az MI-rendszerek használata hátrányos megkülönböztetéshez (diszkriminációhoz) vezethet bizonyos egyének vagy csoportok rovására. Az MI széles körű alkalmazása így korunk egyik legösszetettebb problémáját jelenti nemcsak gazdasági, hanem jogi szempontból is.

Az Európai Unió már 2020-ban felismerte e kihívást, amikor kiadta a mesterséges intelligenciáról szóló Fehér Könyvet.[2] Ebben hangsúlyozták, hogy az MI fejlődése csak akkor nyerheti el a közbizalmat, ha alapvető jogokon és értékeken – így különösen az emberi méltóságon és a magánélet védelmén – alapul. 2021-ben az Európai Bizottság előterjesztette az MI Általános Szabályozásáról szóló, azóta már elfogadott rendelettervezetét (Artificial Intelligence Act),[3] amely egy kockázatalapú megközelítéssel igyekszik orvosolni az MI által jelentett veszélyeket. Világossá teszi, hogy az algoritmusok 'fekete doboz' jellege, komplexitása és adatfüggősége számos alapvető jogot – köztük az egyenlő bánásmódhoz való jogot – veszélyeztethet. Az MI szabályozásának ezért kiemelt célja az esélyegyenlőség magas szintű védelme és a diszkrimináció tilalmának érvényre juttatása.

Jelen tanulmány célja, hogy részletesen feltárja a mesterséges intelligencia és az algoritmikus diszkrimináció jogi kihívásait. A tanulmány a fogalmi keretrendszer bemutatása mellett gyakorlati példákon keresztül mutatja be az algoritmikus diszkrimináció kérdéskörét, és igyekszik hasznos, jövőbe mutató következtetéseket is levonni.

2. FOGALMI KERETEK

2.1. MESTERSÉGES INTELLIGENCIA ÉS ALGORITMIKUS DÖNTÉSHOZATAL

A mesterséges intelligencia tág fogalomkörébe tartoznak mindazon számítógépes rendszerek, amelyek az emberi intelligenciát utánzó műveletek elvégzésére képesek. Idetartoznak a gépi tanulás különböző módszerei (mint például a neurális hálózatok, mélytanulás), a természetesnyelv-feldolgozás, a prediktív analitikai modellek stb. Az MI-rendszerek algoritmusok segítségével, nagy mennyiségű adat feldolgozásával és mintázatok felismerésével hoznak döntéseket vagy tesznek javaslatokat. Algoritmikus döntéshozatalnak nevezzük azt a folyamatot, amikor emberi beavatkozás nélkül, részben vagy teljesen automatizált módon születik meg egy döntés (például hitelbírálat, állásra jelentkező előszűrése, bűnismétlési kockázat becslése stb.) valamely algoritmus kimenetele alapján. Az ilyen rendszerek felhasználása jelentősen átalakítja a hagyományos döntéshozatali mechanizmusokat a magán- és közszférában egyaránt.

Az MI-alapú döntéstámogató vagy döntéshozó eszközök alkalmazása mellett szólnak a hatékonysági és objektivitási várakozások: sokan abban bíznak, hogy a 'gépi döntés' mentes lesz az emberi szubjektivitástól és tévedéstől. A valóságban azonban a technológia sem mentes az emberi tényezőtől. Az algoritmusokat emberek tervezik és tanítják be, mégpedig jellemzően korábbi döntések vagy adatok alapján – így óhatatlanul az emberi döntéshozók bizonyos torzításai, előítéletei és a múlt társadalmi egyenlőtlenségei is átörökítődhetnek a rendszerekbe. Emiatt felmerül, hogy az MI-rendszerek épp az objektivitás látszata révén tehetnek veszélyessé rejtett előítéleteket: az emberek gyakran kritika nélkül fogadják el a gépi ajánlásokat, akár tévesnek tűnő esetekben is, bízva a rendszer vélt csalhatatlanságában.[4] Ez az ún. automatizációs elfogultság jelensége.

* Kutató, Nemzetközi és Uniós Jogi Kutatási Főosztály, Mádl Ferenc Összehasonlító Jogi Intézet.

1 Mezei Kitti: Diszkrimináció az algoritmusok korában. *Magyar Jog*, 2022/6., 331–338.

2 Fehér könyv a mesterséges intelligenciáról – elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:52020DC0065 [letöltve: 2025. 06. 07.]

3 A mesterséges intelligenciáról szóló rendelet – elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX:32024R1689 [letöltve: 2025. 08. 07.]

4 Lukács Adrienn: Digitális diszkrimináció és automatizált döntéshozatal, különös tekintettel a munka világára. *Miskolci Jogi Szemle*, 18. évf., 2. sz., 2023. 103–117. o.

DOI azonosító: https://doi.org/10.51784/FI.2025.1-2.7

2.2. AZ ALGORITMIKUS DISZKRIMINÁCIÓ FOGALMA ÉS FORMÁI

A diszkrimináció jogi értelemben valamely védett tulajdonság (például faj, nem, életkor, vallás, nemzeti vagy etnikai származás, fogyatékosság stb.) alapján történő indokolatlan hátrányos megkülönböztetést jelenti. Az algoritmikus diszkrimináció ennek az a sajátos esete, amikor egy automatizált, algoritmus vezérelte döntéshozatali rendszer eredménye vezet hátrányos különbségtételhez. Lényeges, hogy ilyenkor a megkülönböztetés nem feltétlenül szándékos emberi döntés következménye, hanem az algoritmus működésének 'mellékterméke' lehet. Az MI-rendszerek úgy is generálhatnak diszkriminatív kimeneteket, hogy abban nincs közvetlen emberi szándék vagy akár tudomás. Ha például egy algoritmust olyan adatok alapján tanítottak be, amelyek maguk is tartalmaztak előítéletes vagy egyenlőtlen mintázatokat, akkor a modell ezeket a mintázatokat újratermelheti a jövőbeni döntéseiben.[5]

Az algoritmikus diszkriminációnak két fő formáját különböztetjük meg: direkt (közvetlen) és indirekt (közvetett) diszkriminációt.[6] Direkt diszkriminációról beszélünk, ha az algoritmus kifejezetten egy védett tulajdonság alapján különböztet meg – például ha egy hitelbírálati rendszer nyíltan alacsonyabb pontszámot ad a női jelentkezőknek, vagy egy állás-pályázati szűrőprogram automatikusan kizárja az adott etnikai csoporthoz tartozó jelölteket. Ilyen esetben az algoritmusba szándékosan építenek be olyan szabályt vagy változót, amely a védett csoport hátrányát okozza. Az ilyen közvetlen hátrányos megkülönböztetés egyértelműen sérti az egyenlő bánásmód követelményét, és a legtöbb jogrendszerben tilos.

Indirekt diszkrimináció esetén az algoritmus látszólag semleges módon működik, nem használ explicit módon védett tulajdonságot, a végeredménye mégis aránytalanul hátrányosan érint egy védett csoportot. Ennek fő forrása rendszerint a tanulóadatok torzulása. Ha a betáplált adatok – amelyekből a modell a mintákat megtanulja – már eleve tükrözik a társadalmi egyenlőtlenségeket vagy az emberi döntéshozók korábbi előítéleteit, akkor az algoritmus közvetett módon is diszkriminatív mintázatokat sajátít el. Ilyen esetben a modell nem explicite használja például a nemet vagy a faji hovatartozást döntési kritériumként, de helyettesítő (ún. proxy)[7] változókon keresztül mégis hasonló hatásra jut.[8] Például a jelölt lakcíme, iskolai háttere, szóhasználata stb. alapján juthat olyan következtetésre, amely valójában a jelölt etnikai hátterével vagy nemével függ össze. Az eredmény: bizonyos csoportok – bár az algoritmus formálisan mindenkire ugyanazt a szabályt alkalmazza – szisztematikusan rosszabb elbírálásban részesülnek. Ez a statisztikai torzítás felfogható az emberi jogi értelemben vett közvetett megkülönböztetésnek, amely akkor is tilos lehet, ha első ránézésre semleges eljárásról van szó.

Az algoritmikus diszkrimináció tipikus okai közül néhány: a nem reprezentatív vagy hibás tanulóadatok használata, a múltbeli torz döntéseket tükröző célváltozók választása, a védett jellemzőkkel korreláló proxy-adatok bevitele, valamint a megfelelő emberi felügyelet hiánya a modell működése felett. Az MI-fejlesztés során, ha a történelmi adatok elfogultak, a modell is azzá válik; ha a célként használt 'sikeres döntés' maga is korábbi diszkriminatív gyakorlat eredménye, akkor azt erősíti tovább; ha a rendszer rejtett módon olyan mintázatokat talál, amelyek egy adott csoporthoz tartozást jeleznek, akkor anélkül is hátrányba hozhat egy csoportot, hogy a programozók ezt kifejezetten megadták volna. Mindezek következtében a múlt előítéletei kódolódnak a jövő algoritmusaiba, hacsak tudatos ellenőrzéssel és korrekcióval nem lépünk fel ellenük.

Összefoglalva tehát algoritmikus diszkriminációról akkor beszélünk, amikor egy MI-rendszer előítéletes vagy méltánytalan döntést hoz, amely aránytalanul hátrányosan érint védett csoporthoz tartozó személyeket. Ez történhet direkt módon (kifejezetten a tiltott szempont alapján döntve) vagy indirekt módon (látszólag semleges működésen keresztül). Mindkét eset problémás a jog számára, hiszen az eredmény – az érintett számára – ugyanaz: sérül az egyenlő bánásmód elve, és az illető kevesebb jogot, kedvezményt, lehetőséget kap pusztán egy védett tulajdonsága miatt. Az alábbiakban áttekintjük, a különböző jogrendszerek miként reagálnak erre a kihívásra, de előtte fontos tisztázni az egyenlő bánásmód, az adatvédelem és az emberi jogok általános követelményeit e téren.

3. EGYENLŐ BÁNÁSMÓD, ADATVÉDELEM ÉS EMBERI JOGI SZEMPONTOK

3.1. AZ EGYENLŐ BÁNÁSMÓD KÖVETELMÉNYE ÉS A DISZKRIMINÁCIÓ TILALMA

Az egyenlő bánásmód elve az egyik legalapvetőbb jogi alapelv, mely szerint tilos az indokolatlan megkülönböztetés. Ez az elv megjelenik a nemzetközi egyezményekben, az alkotmányokban és a sarkalatos törvényekben is. Magyarország Alaptörvénye[9] XV. cikkében kimondja, hogy az alapvető jogok mindenkit megilletnek bármiféle megkülönböztetés nélkül, és a törvény előtt mindenki egyenlő. Hasonlóképpen az Európai Unió Alapjogi Chartájának[10] 21. cikke tilt minden fajta diszkriminációt (különösen nem, faji vagy etnikai származás, életkor, fogyatékosság, vallás, szexuális irányultság stb. alapon).

5 *OECD Principles on Artificial Intelligence.* Organisation for Economic Co-operation and Development. – elérhető: https://www.oecd.org/en/topics/ai-principles.html [letöltve: 2025. augusztus 7.]

6 Ilyés Ákos: A gépek lázadása – avagy diszkriminatív algoritmusok az igazságszolgáltatásban?, *Arsboni.hu*, 2022. dec. 14. – elérhető: https://arsboni.hu/a-gepek-lazadasa-avagy-diszkriminativ-algoritmusok-az-igazsagszolgaltatasban/#:~:text=emberi%20logika%20szerinti%20elbírálás%2C%20diszkriminatív,az%2C%20amikor%20szándékosan%20olyan%20prediktív [letöltve: 2025. 08. 07.]

7 X. Wang et.al: *Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices* – elérhető: https://pmc.ncbi.nlm.nih.gov/articles/PMC11148221/?utm_source=chatgpt.com [letöltve: 2025. 08. 07.]

8 Kadocsa Fanni: Workday kontra sokszínűség: miért nem elég, ha nem akartuk?, *HR Power*, 2025. június 18. – elérhető: https://hrpwr.hu/cikk/workday-kontra-sokszinuseg-miert-nem-eleg-ha-nem-akartuk [letöltve: 2025. 08. 07.]

9 Magyarország Alaptörvénye – elérhető: https://net.jogtar.hu/jogszabaly?docid=a1100425.atv [letöltve: 2025. 08. 07.]

10 Európai Unió Alapjogi Charta – elérhető: https://eur-lex.europa.eu/HU/legal-content/summary/charter-of-fundamental-rights-of-the-european-union.html [letöltve: 2025. 08. 07.]

Az Emberi Jogok Európai Egyezménye 14. cikke[11] és Kiegészítő Jegyzőkönyvének 12. cikke is biztosítja a hátrányos megkülönböztetés tilalmát az egyezményben biztosított jogok élvezetében. Ezek a magas szintű elvek együttesen azt rögzítik, hogy senkit nem érhet hátrány pusztán valamely védett tulajdonsága miatt.

A gyakorlatban az egyenlő bánásmód követelménye úgy konkretizálódik, hogy a jog tiltja a közvetlen diszkriminációt (amikor valakivel szemben kifejezetten a védett tulajdonsága miatt bánnak rosszabbul) és a közvetett diszkriminációt is (amikor egy látszólag semleges rendelkezés vagy gyakorlat eredményeként egy védett csoport tagjai aránytalan hátrányba kerülnek). E tilalom alól csak nagyon szűk körben (és szigorú feltételekkel) enged a jog kivételeket – például igazolható, észszerű mérlegelésen alapuló megkülönböztetés formájában, vagy speciális pozitív intézkedések révén.[12]

Az algoritmikus diszkrimináció kihívása éppen az, hogy a meglévő antidiszkriminációs jogszabályok alkalmazhatók-e és hogyan az új technológiai környezetben. Elviekben egy MI által okozott hátrányos megkülönböztetés ugyanúgy sérti az egyenlő bánásmód elvét, mintha azt egy emberi döntéshozó követte volna el. Például ha egy álláspályázókat szűrő algoritmus rendszeresen kiszorítja a fogyatékkal élő vagy kisebbségi jelölteket, akkor ez jogi értelemben felveti a diszkrimináció tilalmának megsértését – függetlenül attól, hogy a döntést egy ember, egy szoftver, vagy a kettő kombinációja hozta. Magyarországon kifejezetten rögzíti is a jog, hogy a hátrányos megkülönböztetés tilos „függetlenül attól, hogy ember vagy algoritmus követte el azt". A 2003. évi CXXV. törvény az egyenlő bánásmódról[13] és esélyegyenlőség előmozdításáról (a továbbiakban: Ebktv.) – hazánk átfogó antidiszkriminációs törvénye – részletesen nevesíti a védett tulajdonságokat és az életviszonyok azon területeit (foglalkoztatás, oktatás, egészségügy, szolgáltatásokhoz való hozzáférés stb.), ahol tilos a diszkrimináció. Az Ebktv. hatálya alá tartozó minden munkáltató, szolgáltató vagy hatóság felelős azért, hogy az általa alkalmazott eljárások – beleértve az esetleges automatizált döntési mechanizmusokat is – ne vezessenek tiltott megkülönböztetéshez. Hasonló szabályok érvényesek az EU-tagállamokban az uniós faji egyenlőségi irányelv (2000/43/EK),[14] foglalkoztatási egyenlőségi irányelv (2000/78/EK), nemek közötti egyenlőség irányelvek (pl. 2006/54/EK)[15] és más jogszabályok alapján, amelyek a nemzeti jog részévé váltak. Az Egyesült Államokban is a polgárjogi törvények (mint a Civil Rights Act VII. címe a munkahelyi diszkrimináció ellen, az ADA a fogyatékosság miatti diszkrimináció ellen, az Age Discrimination in Employment Act[16] az életkor alapján stb.) tiltják, hogy akár a munkáltatók, akár a szolgáltatók diszkrimináljanak – és ez a tiltás kiterjed arra is, ha a döntéshez használt eszköz (pl. egy szoftver) eredménye diszkriminatív.

Mivel az algoritmikus diszkrimináció sokszor közvetett formában jelentkezik, a jogalkalmazásban előtérbe kerül a disparate impact (aránytalan hatás) elemzése. Ez azt vizsgálja, hogy egy adott gyakorlat (pl. egy algoritmus használata) aránytalanul hátrányosan érinti-e a védett csoportba tartozókat másokhoz képest, még ha formálisan semleges is. Az amerikai jogban például a Legfelsőbb Bíróság már az 1970-es évektől elismerte a disparate impact doktrínát, amely alapján bizonyos gyakorlatok akkor is jogsértőnek minősülhetnek, ha nincs bizonyíték szándékos megkülönböztetésre.[17] Ez a megközelítés kulcsfontosságú az MI esetében is, hiszen itt gyakran nincs 'rossz szándékú' diszkriminátor; a hátrány inkább a rendszer működéséből fakadó mellékhatás. Így a jog előtt a sértettnek nem kell feltétlenül szándékot bizonyítania, elég a hatást kimutatnia. Az Európai Unió joga hasonlóképpen tiltja a közvetett diszkriminációt, amelynek megállapításához a hátrányos hatás statisztikai vagy egyéb bizonyítása szükséges.

Összegzésképpen, az egyenlő bánásmód követelménye arra kötelezi mind a kormányzatokat, mind a magánszférát, hogy az általuk használt MI-rendszerek ne eredményezzenek tiltott megkülönböztetést. A kihívás nem annyira az elvi alkalmazhatóságban rejlik – hiszen a meglévő diszkriminációellenes jog elvileg kiterjed ezekre az esetekre is –, hanem inkább a gyakorlati bizonyításban és jogérvényesítésben. Fontos háttérként rögzíteni, hogy az algoritmikus döntéshozatal jogi megítélésében központi szerepe van az egyenlő bánásmód alkotmányos és törvényi követelményének.

3.2. ADATVÉDELMI SZEMPONTOK ÉS AZ AUTOMATIZÁLT DÖNTÉSHOZATAL

Az adatvédelem kérdésköre szorosan összefügg az MI-alapú döntéshozatallal, különösen akkor, ha a döntések személyes adatokon alapulnak. Az Európai Unió Általános Adatvédelmi Rendelete (GDPR)[18] külön rendelkezik az automatizált döntéshozatalról. A GDPR 22. cikke kimondja: az érintett személy jogosult arra, hogy ne terjedjen ki rá egy olyan döntés hatálya, amely kizárólag automatizált adatkezelésen – ideértve a profilalkotást is – alapul, és rá nézve joghatással jár vagy hasonlóképpen jelentős mértékben érinti. Ez a cikk egyfajta védőhálót kíván nyújtani a teljesen gépi döntések ellen,

11 Az Emberi Jogok Európai Egyezménye – elérhető: https://www.echr.coe.int/documents/d/echr/convention_hun [letöltve: 2025. 08. 07.]

12 Weerts Hilde – Xenidis Rapahaele – Tarissan Fabien – Palmer Olsen Henrik – Pechenizkiy Mykola: *Algorithmic Unfairness through the Lens of EU Non-Discrimination Law: Or Why the Law is not a Decision Tree*. ACM Conference on Fairness, Accountability, and Transparency (FAccT, 23). 2023.

13 2003. évi CXXV. törvény az egyenlő bánásmódról – elérhető: https://net.jogtar.hu/jogszabaly?docid=a0300125.tv [letöltve: 2025. 08. 07.]

14 Az uniós faji egyenlőségi irányelv (2000/43/EK) – elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=LEGISSUM:nondiscrimination_principle [letöltve: 2025. 08. 07.]

15 Az Európai Parlament és a Tanács 2006/54/EK irányelve (2006. július 5.) a férfiak és nők közötti esélyegyenlőség és egyenlő bánásmód elvének a foglalkoztatás és munkavégzés területén történő megvalósításáról – elérhető: https://eur-lex.europa.eu/legal-content/HU/ALL/?uri=CELEX:32006L0054 [letöltve: 2025. 08. 07.]

16 Age Discrimination in Employment Act – elérhető: https://www.eeoc.gov/statutes/age-discrimination-employment-act-1967 [letöltve: 2025. 08. 07.]

17 Chiraag Bains: *The legal doctrine that will be key to preventing AI discrimination* – elérhető: https://www.brookings.edu/articles/the-legal-doctrine-that-will-be-key-to-preventing-ai-discrimination/#:~:text=discrimination%20www,without%20having%20to%20prove [letöltve: 2025. 08. 07.]

18 Európai Unió Általános Adatvédelmi Rendelete – elérhető: https://eur-lex.europa.eu/HU/legal-content/summary/general-data-protection-regulation-gdpr.html [letöltve: 2025. 08. 07.]

különösen ha azok hátrányosan érinthetik az egyént (pl. elutasítják egy hitelkérelmét vagy munkára jelentkezését).

A GDPR ugyan enged bizonyos kivételeket (például ha az automatizált döntés szükséges a szerződés teljesítéséhez, vagy az érintett kifejezetten hozzájárult, vagy a tagállami jog megengedi megfelelő garanciák mellett), de ezekben az esetekben is kötelező megfelelő garanciákat alkalmazni. Ilyen garancia különösen az emberi beavatkozás lehetősége: az érintett kérheti, hogy az adott döntést ember vizsgálja felül, vagy hogy legalább megmagyarázzák neki a döntés logikáját. Az automatizált döntések átláthatósága is GDPR-követelmény: a 13–15. cikkek alapján az érintett jogosult tájékoztatást kapni arról, ha rá vonatkozóan automatizált döntéshozatal történik, és ilyenkor információt kérhet az alkalmazott logikáról, valamint arról, hogy a döntés milyen következményekkel jár rá nézve.

Az adatvédelem és a diszkrimináció megelőzése szorosan összefonódik abban az értelemben, hogy a rossz minőségű vagy elfogult adatok használata nemcsak pontatlan eredményekhez vezet, hanem egyes csoportok számára sértő vagy igazságtalan kimenetet is eredményezhet. A GDPR 'pontosság' és 'adattakarékosság' elvei (5. cikk) megkövetelik, hogy az adatkezelés során pontos és releváns adatokra támaszkodjunk. Ha például egy algoritmus olyan személyes adatot használ, amely valójában valamely védett tulajdonságra (pl. faji származásra, egészségi állapotra) utaló proxy, akkor az adatvédelmi szabályok is sérülhetnek. Külön is megemlítendő a különleges adatok kategóriája: a GDPR 9. cikke alapvetően megtiltja a faji vagy etnikai származásra, vallásra, egészségi állapotra stb. vonatkozó adatok kezelését, kivéve szűk kivételek esetén. E tiltás mögött az a cél húzódik, hogy ne lehessen például egy algoritmust közvetlenül a faji hovatartozás alapján 'dönteni tanítani'. Ugyanakkor a gyakorlatban a proxy változók használata miatt ez a tilalom könnyen kijátszhatóvá válhat anélkül, hogy formálisan különleges adatot kezelnének – ezért is van szükség a kimenetek utólagos ellenőrzésére is.

Az adatvédelmi hatóságok is mind nagyobb figyelmet fordítanak az algoritmikus döntéshozatalra. Az elmúlt években több ország adatvédelmi felügyelete vizsgálta, hogy egy-egy MI-rendszer sérti-e a GDPR-t vagy a nemzeti Info törvényt. Hollandia hírhedt esete rávilágított, hogy az adatvédelmi szabályok megszegése és a diszkrimináció együtt járhat: a holland adóhatóság éveken át egy gépi kockázatelemző rendszert használt a családtámogatási visszaélések kiszűrésére, amely jogellenesen gyűjtött és kezelt adatokat (pl. kettős állampolgárságra vonatkozóan), és ennek alapján aránytalanul sok bevándorló hátterű szülőt vádolt meg hibásan csalással. Az eredmény több ezer család tönkretétele lett, és az adatvédelmi hatóság 2022-ben 3,7 millió eurós bírságot szabott ki a jogsértések miatt, kiemelve, hogy a rendszer nem rendelkezett megfelelő jogalappal az adatok ilyen célú használatára, továbbá túl sokáig őrizte az adatokat.[19] Ez az eset is jelzi, hogy az adatvédelmi követelmények (pl. célhoz kötött adatfelhasználás, törlési kötelezettség) megszegése közvetlenül összefügghet a diszkriminatív eredményekkel.

Összességében az adatvédelmi jog arra törekszik, hogy átláthatóbbá és ellenőrizhetőbbé tegye az algoritmikus döntéshozatalt, és biztosítsa, hogy az érintettek jogai (így a tisztességes elbánáshoz való jog) ne sérüljenek. Az adatvédelmi szabályok azonban nem pótolják teljesen a diszkriminációellenes jogot: előfordulhat, hogy egy rendszer formálisan megfelel a GDPR-nek (például beszerezte a hozzájárulást, tájékoztatott az automatizált döntésről stb.), de az eredménye mégis diszkriminatív hatású. Ilyenkor a két terület szabályai párhuzamosan alkalmazandók. A hatékony fellépés érdekében a jogalkotó és a hatóságok igyekeznek összehangoltan kezelni a kérdést – például az Európai Adatvédelmi Testület és az FRA (EU Alapjogi Ügynökség) is több közös állásfoglalást adott ki az MI, a big data és az alapjogok metszéspontjában. Egy jelentés például megállapítja, hogy az algoritmusok torz működése hogyan vezethet diszkriminációhoz – különösen, ha a rendszerek olyan előítéletes adatokon alapulnak, amelyeket a múltbéli emberi döntések vagy strukturális egyenlőtlenségek torzítottak. A jelentés egyértelművé teszi, hogy még a GDPR rendelkezései – például az adatkezelés átláthatósága, az automatizált döntésekről szóló értesítés vagy az érintettek emberi beavatkozáshoz való joga – sem elegendőek önmagukban. Gyakran előfordulhat, hogy egy rendszer formálisan adatvédelmileg 'tiszta', de működésében mégis diszkriminatív hatást eredményez. A dokumentum ezért hangsúlyozza az adatvédelmi és antidiszkriminációs szabályok együttes, összehangolt alkalmazásának szükségességét a valódi jogvédelem érdekében.[20] Az Európai Unió MI rendelete is külön figyelmet fordít az adatok minőségére és a diszkriminációmentesség biztosítására a magas kockázatú MI-rendszereknél, ahogy azt később részletezzük.

3.3. EMBERI JOGI MEGKÖZELÍTÉS

Az algoritmikus döntéshozatal problémáját szélesebb emberi jogi összefüggésben is érdemes szemlélni. Az egyenlő bánásmódhoz való jog – amiről fentebb volt szó – önmagában is alapvető emberi jog. De az MI használata más emberi jogokat is érinthet: ilyen a tisztességes eljáráshoz való jog, a magánélet védelme, az emberi méltóság, sőt a jogorvoslathoz való jog is.

Tisztességes eljárás és átláthatóság: Ha a bíróságok vagy más hatóságok algoritmusok ajánlásaira támaszkodnak (például a büntető igazságszolgáltatásban a vádlott veszélyességének előrejelzésére), felmerül a kérdés, hogy ez mennyire egyeztethető össze a tisztességes eljárás követelményeivel. Az emberi jogok európai egyezményének 6. cikke szerint mindenkinek joga van a tisztességes bírósági eljáráshoz. Egy zárt, megmagyarázhatatlan algoritmus használata a döntéshozatalban aláásható ezt a jogot, hiszen az érintett nem tudhatja, mi alapján született ellene (vagy mellette) a döntés, és így védekezni sem tud hatékonyan. Az Egyesült Államokban a State v. Loomis[21] ügyben

19 *A hollandiai eset beszámolója a sajtóban* (Politico) – elérhető: https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/#:~:text=In%202019%20it%20was%20revealed,spot%20child%20care%20benefits%20fraud [letöltve: 2025. 08. 07.]

20 *Bias in algorithms* – elérhető: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf?utm_source=chatgpt.com [letöltve: 2025. 08. 07.]

21 State v. Loomis: *Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing.* COMMENT ON:881 N.W.2d 749 (Wis. 2016) – elérhető: https://harvardlawreview.org/print/vol-130/state-v-loomis/ [letöltve: 2025. 08. 07.]

a vádlott azzal érvelt, hogy a COMPAS[22] nevű algoritmus használata a büntetéskiszabásnál sérti a tisztességes eljáráshoz való jogát, mivel a védelem nem ismerhette meg a rendszer működését. A Wisconsini Legfelsőbb Bíróság végül engedélyezte a COMPAS[23] eredményeinek figyelembevételét, de feltételeket szabott (pl. nem lehet kizárólag erre alapozni az ítéletet).[24] Európában hasonló dilemmák merülhetnek fel a jövőben a bírósági MI-eszközök kapcsán. Az emberi jogi szempont itt az, hogy az igazságszolgáltatásban (és más közhatalmi döntéseknél) is biztosítani kell az átláthatóságot és indokolhatóságot, mert ezek hiányában sérül az érintett jogorvoslati lehetősége és a hatalom feletti demokratikus kontroll.

Magánélet védelme: Az MI-rendszerek gyakran nagy adattömegeken működnek, és képesek korábban rejtett összefüggéseket feltárni emberek viselkedésében. Ez a magánélethez való jog (privacy) szempontjából aggályos lehet, különösen ha az állam vagy cégek megfigyelésre, profilalkotásra használják. Az algoritmikus diszkrimináció ide is kapcsolódik: például a prediktív rendőrségi algoritmusok azzal a kockázattal járnak, hogy bizonyos etnikai vagy földrajzi közösségeket aránytalanul célpontba vesznek (profilíroznak), így sértve a magánéletüket, és potenciálisan kriminalizálva őket anélkül, hogy egyéni gyanú fennállna. Az emberi jogi keretek – pl. az Európa Tanács adatvédelmi egyezménye (108+ kiegészítés) vagy az ENSZ Polgári és Politikai Jogok Egyezségokmánya (ICCPR) – előírják, hogy a megfigyelés és adatgyűjtés ne vezessen önkényes vagy diszkriminatív gyakorlatokhoz. A magánélet védelme így nemcsak egyéni jog, hanem eszköz is a csoportos diszkrimináció ellen (például ha tiltják bizonyos érzékeny adatok gyűjtését, nehezebb azokat diszkriminációra felhasználni).

Emberi méltóság: Az MI alkalmazása során felmerül, hogy a teljesen gépi döntéshozatal mennyiben tárgyiasítja az embert. Az EU AI rendelete[25] is abból indul ki, hogy az MI-nek emberközpontúnak kell maradnia, tiszteletben tartva az emberi méltóságot. Az emberi méltóságot sértheti, ha valakit pusztán statisztikai alapon, egy algoritmus számként kezel, és például megfoszt egy lehetőségtől anélkül, hogy ügyének egyedi körülményeit bárki tekintetbe venné. Az ENSZ 2021-ben elfogadott MI Etikai Ajánlása szintén hangsúlyozza a diszkrimináció tilalmát és a méltóság védelmét az MI használatában.[26] Az emberi beavatkozás lehetősége etikai és jogi szempontból is fontos: végső soron az ember felelőssége kell maradjon a döntés, nem lehet azt teljesen a gépre hárítani, mert ezzel a jog elszámoltathatósági mechanizmusai (felelősségre vonás, jogorvoslat) kiüresednének.

Jogorvoslat és felelősségre vonás: Minden jogállamban alapelv, hogy ha jogsérelem ér valakit, akkor hatékony jogorvoslat illeti meg. Az algoritmikus diszkriminációnál komoly gondot okozhat ennek biztosítása. Egyrészt, az áldozatok sokszor nem is tudnak arról, hogy őket diszkrimináció érte – hiszen lehet, hogy fogalmuk sincs róla, hogy egy algoritmus miatt utasították el őket állásról vagy adtak nekik rosszabb feltételekkel hitelt. A transzparencia hiánya miatt a sértés rejtve maradhat. Másrészt, ha tudnak is róla, nehéz bizonyítani, hogy a hátrány oka éppen a tiltott megkülönböztetés volt, és nem valami legitim tényező. Harmadrészt, felmerül a kérdés, hogy kit lehet perelni vagy felelősségre vonni: a szoftver fejlesztőjét, az üzemeltetőt, az adatot szolgáltatót, vagy mindegyiket? Ezekre a dilemmákra a jogrendszerek jelenleg keresik a választ. Emberi jogi szempontból mindenesetre elvárás, hogy az állam olyan keretet teremtsen, amelyben van felelőse a jogsértő algoritmikus döntésnek és van lehetőség a sérelem orvoslására. Ezt az igényt tükrözi például az AI Act is, amely kimondja, hogy a részes feleknek biztosítaniuk kell a hatékony jogorvoslatot az MI-rendszerek által okozott jogsértések esetére. Hasonlóképpen az EU kidolgozott egy MI felelősségi irányelvjavaslatot,[27] amely a polgári jogi kárfelelősség terén könnyítené meg a bizonyítást az MI okozta károk esetén.

Összefoglalva, a mesterséges intelligencia alkalmazását övező egyik legfontosabb emberi jogi követelmény, hogy az ember maradjon az úr – azaz az MI ne csorbíthassa az ember alapvető jogait, se közvetlen diszkriminációval, se azzal, hogy kikerüli az emberi kontrollt. Az egyenlő méltóság és egyenlő jogok elve minden modern jogrendszer sarokköve; ennek az algoritmusok korában is érvényesülnie kell.

4. ESETTANULMÁNYOK

A következőkben néhány konkrét példán keresztül mutatom be, miként jelenik meg a gyakorlatban az algoritmikus diszkrimináció, és milyen jogi reakciók követték az egyes eseteket.

4.1. COMPAS ÉS A BŰNISMÉTLÉSI KOCKÁZAT BECSLÉSE (USA)

A Correctional Offender Management Profiling for Alternative Sanctions, röviden COMPAS nevű algoritmust az Egyesült Államok egyes tagállamaiban használják annak előrejelzésére, hogy az őrizetbe vettek közül ki milyen eséllyel fog újra bűncselekményt elkövetni. A rendszert a bíróságok a vádlott kockázati besorolására alkalmazzák (pl. feltételes szabadlábra bocsátásnál vagy ítéletnél hivatkozási pontként). 2016-ban azonban egy oknyomozó elemzés rámutatott, hogy a COMPAS előrejelzései faji elfogultságot hordoznak: a fekete vádlottakat a rendszer sokkal nagyobb arányban sorolta be tévesen magas kockázatúnak,

22 A COMPAS mozaikszó a következőt jelenti: Correctional Offender Management Profiling for Alternative Sanctions, vagyis 'büntetés-végrehajtási elkövetőkezelési profilalkotás alternatív szankciókhoz'.

23 Tim Brennan: Northpointe Institute for Public Management Inc., Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System, *36 Criminal Justice and Behavior 21*, 2009.

24 Mezei i. m. 331–338. o.

25 Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (1) bekezdés.

26 UNESCO Ajánlás a mesterséges intelligencia etikájáról – elérhető: https://unesco.hu/data/UNESCO_Ajanlas_mesterseges_intelligencia.pdf [letöltve: 2025. 08. 07.]

27 Az Európai Parlament és a Tanács a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv (javaslat) – elérhető: https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:52022PC0496 [letöltve: 2025. 08. 07.]

mint a fehéreket.[28] A statisztikák szerint a fekete vádlottaknál csak 20% eséllyel tévedett pozitív irányba (azaz ők veszélytelenek voltak, mégis magas kockázatot kaptak), míg a fehéreknél 47% eséllyel kapott valaki alacsony kockázati pontszámot annak ellenére, hogy valójában visszaesett volna.[29] Ezek az adatok hatalmas vitát váltottak ki. Jogilag a kérdés a Loomis-ügyben került felszínre (Wisconsin, 2017), ahol a vádlott azt állította, hogy a COMPAS használata megsértette az alkotmányos tisztességes eljáráshoz való jogát, és közvetetten faji diszkriminációhoz vezet. A Wisconsini Legfelsőbb Bíróság végül úgy döntött, hogy a COMPAS eredménye használható, de a bírónak figyelmeztetésekkel ellátva kell kezelnie: nem alapozhat kizárólag erre, és tudomásul kell venni a rendszer korlátait. Ez egy kompromisszumos ítélet volt. Közben a közfelháborodás nyomán Kalifornia és más államok elkezdték felülvizsgálni a prediktív algoritmusaik alkalmazását. A COMPAS esete rávilágított arra, hogy bár látszólag minden vádlott kap egy 'egyéni' pontszámot, a rendszer mégis rendszerszinten diszkriminatív tud lenni. Az USA-ban azóta a tudományos és jogi diskurzus központi témája lett, hogyan kellene auditálni és korrigálni az ilyen igazságszolgáltatási MI-rendszereket, vagy egyáltalán használni kell-e őket.

Prediktív rendőrség és profilalkotás: Számos ország rendőrsége kísérletezik ún. prediktív policing programokkal, amelyek adatokat (korábbi bűncselekmények helye, időpontja, személyek profiljai) elemezve igyekeznek megjósolni, hol és ki fog nagy valószínűséggel bűncselekményt elkövetni. Ilyen rendszerek működnek például az USA több városában (PredPol néven is ismert szoftverek),[30] de Európában is (pl. az Egyesült Királyságban[31] és Németországban[32] pilot projektek). A veszély abban rejlik, hogy ezek a programok gyakran a múltbeli rendőri adatokra épülnek, amelyek már tartalmazhatnak elfogultságot – például egy bizonyos negyedből több letartóztatási adat van (lehet, hogy azért, mert ott intenzívebb a rendőri jelenlét, nem mert több a bűnöző), így a rendszer azt fogja 'jósolni', hogy ott a jövőben is több bűntény lesz.[33] Ennek eredményeként az adott (mondjuk etnikai kisebbség lakta) negyedben még több rendőr járőrözik, még több embert állítanak meg, és öngerjesztő módon be is igazolják a 'jóslatot'. Ez a fajta algoritmikus 'vörös csőr effektus' (feedback loop) egyértelműen diszkriminatív hatású lehet, noha a program sehol nem használ faji adatot – csupán a földrajzi, bűnügyi statisztikát.[34] Hollandiában egy időben használtak egy profilkészítő algoritmust a rendőrségnél, ami a 'problémás' fiatalokat próbálta kiszűrni iskolai adatok alapján, és utólag derült ki, hogy a bevándorló hátterű fiúk felé torzult. Miután ez kiderült, a rendszert leállították.[35] Az Egyesült Királyságban 2019-ben Birmingham város tanácsa moratóriumot hirdetett az új prediktív rendészeti szoftverek bevezetésére, amíg azoknál nem bizonyított a bias-mentesség.[36] Itt lép be az emberi jogi megközelítés: a rendőrség köteles a törvény előtti egyenlőséget és az ECHR 14. cikket[37] betartani. Ha egy MI-eszköz azt eredményezi, hogy pl. a fekete lakosságot aránytalanul zaklatják 'igazoltatás és motozás' (stop and frisk) keretében, az sérti a diszkrimináció tilalmát, akkor is, ha formálisan a szoftver csak bűnözési kockázatot jelölt. Az Európa Tanács MI egyezménye kimondja,[38] hogy a migráció- és határellenőrzés terén használt MI-rendszereknél is figyelni kell a diszkriminációmentességre – ez reflektál arra, hogy pl. hazugságvizsgáló MI-t teszteltek menedékkérőknél az EU-ban, ami rendkívül vitatott lett etikailag.[39]

4.2. AZ AMAZON ÖNÉLETRAJZSZŰRŐ BOTRÁNYA

Az egyik legismertebb példa a mesterséges intelligencia alkalmazásával kapcsolatos diszkriminációs kockázatokra a 2018-as Amazon-eset, amely széles körű sajtó- és szakmai visszhangot váltott ki. A cég belső kísérleti projekt keretében fejlesztett egy MI-rendszert, amely a beérkező önéletrajzokat rangsorolta informatikai és mérnöki pozíciókra. Az algoritmust korábbi felvételi adatokon tanították, azonban ezek az adatok jelentős nemi torzítást hordoztak: a múltban az Amazon technikai pozícióit túlnyomórészt férfiak töltötték be.[40] Ennek következtében az MI-rendszer megtanulta, hogy a férfi jelöltek jellemzői preferáltak – például visszasorolta azokat az önéletrajzokat, amelyekben szerepelt a 'women's' szó (mint például 'women's chess club captain'), vagy olyan kifejezéseket, amelyek női iskolai háttérre utaltak.[41] A modell ezzel rejtett nemi alapú diszkriminációt valósított

28 Borgesius, Frederik Zuiderveen: *Discrimination, artificial intelligence, and algorithmic-decision making*. Council of Europe, 2018. 23–25.

29 Borgesius i. m. 23–25.; valamint lásd erről részletesen Dieterich, William–Mendoza, Christina–Brennan, Tim: *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*. Northpointe, 2016.; Angwin, Julia et al.: Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *ProPublica,* 2016.; Larson, Jeff et al.: How We Analyzed the COMPAS Recidivism Algorithm. ProPublica, 2016.

30 Andrew Guthrie Ferguson: The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. *NYU Press*. 2017.

31 Amnesty International UK (2020). Trapped in the Matrix: Secrecy, stigma, and bias in predictive policing. https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf [letöltve: 2025. 08. 07.]

32 Simon Egbert: Predictive Policing in Germany: Between Precaution and Pre-emption. *European Journal of Risk Regulation*, 10(3), 2019. 439–452.

33 Solon Barocas – Andrew D. Selbst.: Big Data's Disparate Impact. *California Law Review*, 104(3), 2016. 671–732.

34 Virginia Eubanks: Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. *St. Martin's Press*. 2018.

35 Mutsaers, P.; Nuenen, T. van 2023, Predictively policed: The Dutch CAS case and its forerunners, Part of book or chapter of book (Beek, J.; Bierschenk, T.; Kolloch, A. (ed.), Policing race, ethnicity and culture: Ethnographic perspectives from Europe, pp. 72-94) – elérhető [online: https://repository.ubn.ru.nl/bitstream/handle/2066/289967/289967.pdf?sequence=1&isAllowed=y [letöltve: 2025. november 10.]

36 Birmingham City Council (2019). Moratorium on Predictive Policing Tools – Council Decision.

37 Európai Emberi Jogi Egyezmény, 14. cikk – elérhető: https://www.echr.coe.int/documents/convention_eng.pdf [letöltve: 2025. augusztus 7.]

38 Council of Europe (2024). Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law. – elérhető: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence [letöltve: 2025. augusztus 7.]

39 FRA (2022). Fundamental rights and the use of artificial intelligence in migration. – elérhető: https://fra.europa.eu/en/publication/2022/fundamental-rights-report-2022 [letöltve: 2025. augusztus 7.]

40 Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women.* Reuters. – elérhető: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G [letöltve: 2025. augusztus 7.]

41 Inioluwa Deborah Raji – Joy Buolamwini.: Actionable Auditing: *Investigating the Impact of Publicly Naming Biased Performance Results of*

meg, noha explicit nemi adatokat nem használt. A projektet belső vizsgálat után 2018-ban leállították, még mielőtt éles működésbe állt volna.[42] Bár az Amazon ellen nem indult hivatalos jogi eljárás – mivel a rendszer nem került nyilvános alkalmazásra és nem volt konkrét panaszos – az eset rávilágított az AI-rendszerek 'tanulási torzításainak' veszélyeire. Azóta több HR-technológiával foglalkozó cég és nagyvállalat bevezetett független auditálási folyamatokat a mesterséges intelligenciát használó toborzási eszközeire, hogy kiszűrje az elfogultságokat.[43] Az Amerikai Egyenlő Foglalkoztatási Hivatal (EEOC) külön tájékoztató kampányt indított, amelyben felhívja a figyelmet arra, hogy a munkaadók kötelesek megbizonyosodni arról, hogy az általuk használt automatizált eszközök nem diszkriminálnak.[44] Az Európai Unió mesterséges intelligencia rendelete (AI Act) e toborzási célú MI-rendszereket a 'magas kockázatú' kategóriába sorolja, így szigorúbb jogi követelmények vonatkoznak majd rájuk – különösen a torzítások kimutatására és kockázatcsökkentésre vonatkozó előírások.

Magyarországon egyelőre kevesebb az MI-alapú HR-rendszer, de már megjelentek egyszerűbb tesztelő és értékelő megoldások, például automatikus nyelvi elemzést alkalmazó rendszerek. A tanulság azonban globális: mivel a munkaerőpiac történelmileg torz társadalmi struktúrákra épül, az algoritmusok könnyen ezeket a torzulásokat vehetik alapul – továbbörökítve a meglévő egyenlőtlenségeket.[45]

4.3. HITELBÍRÁLATI ALGORITMUSOK ÉS FAJI EGYENLŐTLENSÉG

A bankok és fintech cégek egyre szélesebb körben alkalmaznak mesterséges intelligencián alapuló credit scoring rendszereket a hitelképesség értékelésére. Bár az Egyesült Államokban az Equal Credit Opportunity Act (ECOA) tiltja a nemi, faji vagy egyéb védett tulajdonságon alapuló diszkriminációt a hitelezésben, a modern algoritmusok jellemzően 'black box' jellegűek, és számos, közvetett módon torzító tényezőt tartalmazhatnak.[46] 2020-ban az Apple és a Goldman Sachs közös terméke, az Apple Card került reflektorfénybe, miután több felhasználó – köztük az Apple társalapítója, Steve Wozniak – nyilvánosan jelezte, hogy a női igénylők azonos pénzügyi háttérrel is lényegesen alacsonyabb hitelkeretet kaptak, mint férfi társaik. Egy esetben egy házaspár női tagja – noha jobb egyéni hitelmutatókkal rendelkezett – tízszer alacsonyabb keretet kapott, mint férje. A New York-i pénzügyi felügyelet (DFS) vizsgálatot indított, ám végül nem állapított meg szándékos jogsértést, részben mert a modell működése nem volt kellően átlátható.[47]

Ez az eset jól példázza, milyen jogi és bizonyítási nehézségeket vet fel az algoritmikus döntéshozatal transzparenciahiánya: noha az Apple Card modell formálisan nem használt nemi adatot, a végeredmény mégis jelentős hátrányt okozott nők számára. Az érintettek nehezen tudták igazolni, hogy hátrányos megkülönböztetés történt.

Európában a GDPR 22. cikke biztosítja az érintettek számára az emberi beavatkozás jogát automatizált döntések esetén, továbbá egyes országok (pl. Franciaország) jogszabályai tiltják bizonyos érzékeny adatok (pl. nem, lakcím) figyelembevételét hitelbírálat során.[48] Ennek ellenére kutatások szerint például Németországban a SCHUFA által alkalmazott hitelmodellek a migrációs háttérrel rendelkező személyeket hátrányosan érinthetik – vélhetően lakóhely vagy más közvetett változók miatt. Civil nyomás hatására 2021-ben a SCHUFA egyes algoritmusait részben nyilvánosságra kellett hozni, és a német politika célul tűzte ki az algoritmikus pénzügyi döntések szigorúbb felügyeletét.

5. KÖVETKEZTETÉSEK

A mesterséges intelligencia rohamos fejlődése új dimenzióba helyezte a diszkrimináció tilalmának érvényesítését. Az algoritmikus döntéshozatal – legyen szó közszolgáltatásokról, igazságszolgáltatásról vagy piaci szolgáltatásokról – a hatékonyság és objektivitás ígéretével érkezett, de világossá vált, hogy emberi felügyelet és megfelelő jogi kontroll nélkül a technológia a meglévő társadalmi egyenlőtlenségek felnagyításának eszközévé válhat.

A mesterséges intelligencia nem csupán technológiai, hanem társadalmi innováció is, amely csak akkor tekinthető sikeresnek, ha mindenki javát szolgálja, nem csak egyes kiváltságos csoportokét. A jog feladata, hogy közvetítse az alapvető értékeket ebbe az új környezetbe is. Az egyenlő bánásmód, az emberi méltóság tisztelete, az átláthatóság és elszámoltathatóság olyan elvek, amelyekből nem engedhetünk akkor sem, ha a döntéshozó szerepét részben gépek veszik át.

A tanulmányban bemutatott esetek és jogi reakciók azt igazolják, hogy az algoritmikus diszkrimináció valós és sürgető probléma, de megfelelő stratégiával és szabályozással kezelhető. A megoldás nem csupán a jogalkotó kezében van: szükség van a technológiai szféra felelősségvállalására, a tudományos szféra inputjára és a civil kontrollra is. Fontos, hogy a mesterséges intelligencia úgy integrálódjon társadalmunkba, hogy közben ne csorbuljanak az egyenlő emberi jogok. Ez kihívásokkal teli feladat, de nem megoldhatatlan: a jog evolúciója mindig is követte a technológia fejlődését, és jelen esetben sincs ez másként.

Commercial AI Products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019.

42 Ziad Obermeyer – Brian Powers – Christine Vogeli – Sendhil Mullainathan: Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 2019. 447–453.

43 Solon Barocas – Moritz Hardt – Arvind Narayanan: Fairness and Machine Learning. fairmlbook.org. 2019. – elérhető: https://fairmlbook.org/ [letöltve: 2025. augusztus 7.]

44 U.S. Equal Employment Opportunity Commission (EEOC). (2023). Artificial Intelligence and Algorithmic Fairness in Employment. – elérhető: https://www.eeoc.gov/ [letöltve: 2025. augusztus 7.]

45 Eubanks i. m.

46 Barocas–Selbst i. m. 671–732. o.

47 New York State Department of Financial Services (NY DFS). (2021). Report on Apple Card Investigation. – elérhető: https://www.dfs.ny.gov [letöltve: 2025. 08. 07.]

48 Sandra Wachter – Brent Mittelstadt – Luciano Floridi: Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 2019. 76–99. o.

Konklúzióként megállapítható, hogy az MI és az algoritmikus diszkrimináció viszonyának rendezése folyamatban lévő, dinamikus terület. A jogi kihívásokra adott válaszok egyre határozottabbak: a nemzetközi és nemzeti jogalkotó közösség is felismerte, hogy a jövő digitális társadalmában az algoritmusoknak is alá kell vetniük magukat a jog uralmának. A törvény előtti egyenlőség elvének az adatvezérelt világban is érvényesülnie kell – ez közös felelősségünk és érdekünk. A megfelelő jogi keretek kialakításával, az ellenőrzési mechanizmusok kiépítésével és az össztársadalmi odafigyeléssel elérhetjük, hogy a mesterséges intelligencia ne az előítéletek konzerválásának, hanem a méltányosság és haladás előmozdításának eszköze legyen.