

MIKE NIMRÓD

Előítélet, diszkrimináció és gazdasági egyenlőtlenségek

A mesterséges intelligencia etikai kihívásai

TANULMÁNY

DOI: 10.56699/MT.2024.4.5

ABSZTRAKT

Az MI-technológiák gyors fejlődése és elterjedése számos etikai kérdést vet fel, amelyek alapos vizsgálatot és szabályozást igényelnek. Jelen tanulmány célja az MI alkalmazásával kapcsolatos legfontosabb társadalmi és etikai aggályok feltérképezése és ismertetése. A tanulmány az MI-technológiák etikai dilemmáit két fő területre osztja: előítéletesség és diszkrimináció, valamint munkavállalók áthelyezése és gazdasági egyenlőtlenségek. Az MI-rendszerek gyakran továbbviszik a tanító adatokban meglévő torzításokat, ami igazságtalansághoz vezethet a társadalom különböző csoportjai között. Emellett az MI automatizálása jelentős munkahelyvesztést és jövedelmi egyenlőtlenségeket eredményez, mivel a technológiai előnyöket főként a nagyvállalatok és a magasan képzett szakemberek élvezik. A tanulmány számos példát mutat be az MI alkalmazásának etikai kihívásaira, és javaslatot tesz a szabályozási keretek és etikai normák kidolgozására.

KULCSSZAVAK: mesterséges intelligencia, etikai dilemmák, előítélet, diszkrimináció, gazdasági egyenlőtlenségek

A mesterséges intelligencia (a továbbiakban: MI) fejlesztésével párhuzamosan egyre szélesebb körűvé válnak az alkalmazási területei, valamint az ezzel a fejlett technológiai eszközzel kapcsolatos etikai kérdések is. Az MI különböző etikai irányelvek kidolgozásához vezetett a pozitív alkalmazások előmozdítása és a negatív következmények minimalizálása érdekében. Számos új MI-technológiát építettek be a világ mindennapi működésébe, amilyenek például az egészségügyben használt MI-alapú eszközök, az MI által támogatott autók vagy az arcfelismerő rendszerek az üzleti életben és a szolgáltatóiparban. Kétségtelen, hogy az elmúlt évek egyik legszembevetőbb vívmánya a ChatGPT nevű alkalmazás volt.

A sikeres alkalmazások ellenére különböző etikai kérdések merültek fel, amelyek átfogó felmérést és elemzést igényelnek. Az MI etikájának kérdése még mindig jelentősen alulkiértékelve van, ezért szükség van a benne rejlő etikai potenciál feltárására, hogy a fejlesztők és a társadalom egésze számára megfontolandó etikai koncepciót dolgozzunk ki. Jelen tanulmány az MI-vel kapcsolatos etikai kérdések meghatározása érdekében széles körű szakirodalmat tekint át. Célja az MI alkalmazásával kapcsolatos legfontosabb társadalmi és etikai aggályok feltérképezése és ismertetése.

A tanulmány így épül fel: először az alapvető etikai dilemmák beazonosítására kerül sor, majd az előítéletesség és diszkrimináció kérdéskörének mélyebb elemzését olvashatjuk. Ezt követően az írás kitér a munkavállalók áthelyezésére és a gazdasági egyenlőtlenségre mint etikai dilemmákra. Az értekezés konklúzióval zárul.

Az előítéletesség, diszkrimináció, a munkavállalók áthelyezése és a gazdasági egyenlőtlenségek olyan etikai dilemmákat vetnek fel, amelyek már az MI-technológiák megjelenése előtt is léteztek, de amelyeket az MI-technológiák alkalmazása tovább fokozhat. A tanulmány ezek vizsgálatára fókuszál. Egyéb kérdéseket, mint például az adatvédelmi és adatbiztonsági aggályok, a terjedelmi korlátok miatt nem tárgyalunk. A tanulmányban az „etikai dilemmák” és „etikai kérdések” kifejezések szinonimaként értelmezendők.

ETIKAI DILEMMÁK AZ MI TERÜLETÉN

Az etikai elvek azok az erkölcsi iránymutatások, amelyek az egyének vagy csoportok cselekedeteit és viselkedését irányítják. Ezek az elvek segítenek meghatározni, mi minősül „adekvát” vagy „helytelen” viselkedésnek egy adott élethelyzetben, és magukban foglalják az egyének vagy csoportok erkölcsi kötelelességeit is.

Az MI területén az etikai keretek kidolgozása különösen fontos, tekintettel a technológia feltörekvő és újszerű jellegére. Mark Coeckelbergh (2020) rámutat, hogy mivel az MI különböző interakciókat folytat az emberekkel, elengedhetetlenül szükséges az MI-alapú fejlesztések és technológiai alkalmazások etikai kereteinek kidolgozása. A szakirodalom arra is rámutat, hogy az MI használatának etikai dilemmái gyakran a más fejlett technológiákkal való kombinációból erednek. Keng Siau és Weiyu Wang (2020) szerint az MI etikáját még mindig nem veszik kellőképpen figyelembe, mivel a legtöbb MI-alapú technológia programozása korlátozza autonóm alkalmazásukat. Ugyanakkor a mindennapokban már ezen a téren is változás figyelhető meg.

Az MI-alkalmazások az emberi élet számos területét befolyásolják, ezért egyértelműen szükség van szabályozásra és átfogó felülvizsgálatra. Erre vonatkozóan több szabályozási törekvés is létezik. Lássunk néhány példát:

a) Az Egyesült Államokban a 2023. október 30-án kiadott „Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence” (White House, 2023) célja az MI-rendszerek biztonságának és megbízhatóságának növelése, valamint az amerikaiak magánéletének védelme.

b) Kínában a „New Generation Artificial Intelligence Development Plan” (State Council of PRC, 2017) irányozza elő az MI fejlesztését és alkalmazását, beleértve az etikai normák és szabványok kidolgozását a társadalmi stabilitás és a nemzetbiztonság megőrzése érdekében.

c) Az Európai Unióban az „Artificial Intelligence Act” (EUR-Lex, 2021) javasolja az MI-rendszerek kockázatalapú szabályozását, különös hangsúlyt fektetve a magas kockázatú alkalmazásokra és az etikai elvek betartására, hogy biztosítsák az alapvető jogok védelmét.

Összességében mindhárom szabályozás célja, hogy kezelje az MI-technológiák által felvetett alapvető etikai dilemmákat. Ezek a dilemmák a következő két kategóriába sorolhatók: az MI fejlesztését érintő, valamint az MI-alapú rendszerek használata által okozott etikai dilemmák.

Az MI fejlesztését érintő néhány etikai dilemmát az olyan tényezők okozzák, mint az MI-rendszereket betanító adathalmazokban megjelenő torzítás (*data bias*), az adatbiztonság (*data security*) és a személyes adatok védelme (*data protection*). Például egy olyan MI-alapú alkalmazásnak, mint a ChatGPT, a szakirodalom szerint erőforrás-igényes és erősen környezetszennyező a fenntartása és működtetése (Hale, 2023). Matthew Elmore (2023) pedig rávilágít, hogy specifikus etikai dilemma a ChatGPT esetén, hogy az OpenAI kizsákmányolt és alulfizetett adatszolgáltatókra támaszkodott afrikai és más alacsony jövedelmű országokban.

Az MI-alapú alkalmazások használatából eredő etikai dilemmákat okoz az előítéletesség és diszkrimináció (*bias and discrimination*), a munkahelyek kiszorítása vagy a munkavállalók áthelyezése (*job displacement*) és a gazdasági egyenlőtlenségek (*economic inequity*). A továbbiakban a tanulmány e dilemmákat kívánja röviden körüljárni.

ELŐÍTÉLETESÉG ÉS DISZKRIMINÁCIÓ

Az MI-alapú technológiákat emberek fejlesztik és tanítják, s az általuk létrehozott adatbevitel révén kapnak e technológiák tanító adatokat. Ennek következtében a tanító adatok, amelyeket a rendszerekbe táplálnak, elfogultságot és diszkriminációt eredményezhetnek. Az MI-alapú eszközök képesek olyan hatalmas mennyiségű és összetettségű adatot gyűjteni és értékelni, amelynek a feldolgozása meghaladja az emberi képességeket, s így ezek beépíthetők az elemzésekbe és a döntéshozatali folyamatokba (Byrne, 2021). Ennek következményeként különböző (például bünyügyi és igazságszolgáltatási, toborzási, valamint monitoring- és értékelési) rendszerekben fel lehetett fedezni az MI elfogultságát és diszkriminatív megnyilvánulásait. Ez a tény különösen akkor aggodalomkeltő, ha figyelembe vesszük, hogy az ilyen döntéstámogatási algoritmusoknak a kifejlesztésekor épp azt tűzték ki célul, hogy azok a különféle rendszerekben megjelenő elfogultságot csökkentsék, és megkönnyítsék a döntést. Ehhez képest a gyakorlatban megjelenő MI-alkalmazások ennek épp az ellenkezőjét bizonyították. Konkrétan azt, hogy nem tudják teljesen megszüntetni a diszkriminációt. Ennek egy példája az Amazon nevű

vállalati csoporthoz köthető, ahol MI-alapú toborzási rendszert fejlesztettek ki, hogy automatizálja az álláspályázatok átvizsgálását, és azonosítsa a legjobb jelölteket. Az algoritmus azonban a férfiakat preferálta, mivel az előző tíz évben benyújtott pályázatokból tanult, amelyek nagyrészt férfiakról származtak. Így hátrányosan megkülönböztette a női jelentkezőket, például alacsonyabb pontszámot adott azoknak a pályázatoknak, amelyek tartalmazták például a „női” szót („női röplabdaklub”) vagy olyan egyetemek nevét, ahol több női hallgató volt (Dastin, 2018).

Az MI-alapú alkalmazások jelentős aggodalmakat vetettek fel és potenciális kárt okoztak a büntető- és igazságszolgáltatási rendszerekben is. Az MI elfogultsága a büntetőrendszerben a prediktív bűnüldözési algoritmusok révén mutatkozott meg, amelyekben az MI-t a bűnügyi adatok elemzésére, illetve a bűncselekmények valószínűségének és helyének előrejelzésére használták (Berk, 2021). A bűnüldözési rendszerből eredő elfogultság a szakirodalom szerint diszkriminációhoz és egyenlőtlenségekhez vezetett a büntető igazságszolgáltatás terén.

Az Amerikai Egyesült Államokban elhíresült példák jól szemléltetik ezt a fajta elfogultságot. Elsőként ismertetendő a PredPol: ez egy bűnügyi előre jelző rendszer, amelyet több rendőrkapitányság használt az Egyesült Államokban (Angwin et al., 2016). A történelmi bűnügyi adatok alapján próbálta előre jelezni a jövőbeni bűncselekmények helyét és időpontját. Mivel azonban a történelmi bűnügyi adatok magukban foglalták a rendőrség általános elfogultságát, a rendszer hajlamos volt a már korábban is túlzottan megfigyelt és járőrözött kisebbségi közösségeket megcélolni (Lum–Isaac, 2016). Ennek eredményeként e közösségek még több rendőri jelenléte tapasztaltak, ami tovább növelte az igazságtalanságot és a diszkriminációt.

A második példa a COMPAS nevű kockázatbecslő rendszer, amelyet szintén az Egyesült Államokban használnak, hogy értékeljék a bűnelkövetők újabb bűncselekmény-elkövetésének valószínűségét (Angwin et al., 2016). Egy vizsgálat kimutatta, hogy a rendszer faji elfogultságot mutatott: az afroamerikai elkövetők esetében nagyobb valószínűséggel jelzett magasabb kockázatot, még akkor is, ha azonos tényezőket vett figyelembe a fehér bőrű elkövetőkkel szemben (Dressel–Farid, 2018). Ez részben annak volt köszönhető, hogy a rendszer tanító adatbázisában már meglévő faji elfogultságok szerepeltek.

Ezek a példák jól illusztrálják, hogy az MI-alapú rendszerek mennyire függenek a tanító adatkészlet minőségétől, és mennyire fontos, hogy kiegyensúlyozott, reprezentatív és elfogulatlan adatbázisokat használjanak az ilyen rendszerek fejlesztése során. Itt fontos hangsúlyozni, hogy az EU-ban az MI szabályozására készült rendelet – a már említett Artificial Intelligence Act – tiltott alkalmazási módszernek nevezi a profilalkotásra épülő prediktív rendészeti eljárásra használni kívánt MI-alapú rendszereket (EUR-Lex, 2021).

Emellett lényeges szempont, amint a fenti példák is mutatják, hogy az MI-alapú algoritmusok történelmi adatokra alapozzák működésüket. Az algoritmusrendszerek frissítése nélkül az MI potenciális fenyegetést jelenthet, ha a korábbi adatok elfogultságot és diszkriminatív gyakorlatot tartalmaznak.

Xavier Ferrer és munkatársai (2021) szerint mivel az MI adatait emberek táplálják be, a torzítás nemcsak a kezdeti betáplált adatokból származhat, hanem a fejlesztők nemtörődömségének, hanyagságának eredménye is lehet. Így az elfogultság mint a torzítás nem kívánt következménye nem csupán a rossz minőségű tréningadatok használata miatt jöhet létre, hanem szándékkal, vagy gondatlan algoritmustervezés útján is. Egy másik emberi tényező az, hogy a fejlesztők elmulasztják az információk diverzifikálását. Ez magában foglalja különböző háttérű, társadalmi osztályú, perspektívájú és tapasztalatú személyek kihagyását vagy előnyben részesítését, ami szintén diszkriminatív eredményekhez vezethet (Chen, 2023).

Az MI-eszközök elfogultsága és diszkriminációja számos etikai és jogi problémát okozott, beleértve a korlátozott elszámoltathatóságot és átláthatóságot, a kisebbségi közösségekre gyakorolt hatást és a nem szándékos elfogultságokat, amelyek társadalmi és gazdasági egyenlőtlenségekhez vezetnek. Így egyfajta ajánlásként megfogalmazható, hogy az EU-ban kialakult álláspont alapvetően ennek az etikai problémának a feloldását hivatott megtestesíteni: az MI-alapú technológiákat nem szabad túlzott mértékben döntéshozatalra használni, sokkal inkább a döntéstámogatás lehet az etikus cél. Az MI-rendszerek igazságos és nem diszkriminatív módon történő fejlesztése szintén csökkentheti az elfogultságot és diszkriminációt, s javíthatja az eredményeket.

Az MI-rendszerek túlzott mértékű használata (*overreliance*) jelentős kockázatokat hordoz magában, különösen akkor, ha a felhasználók kritikus területeken is ellenőrzés nélkül hagyatkoznak az MI válaszára. Az *overreliance* problémája azt jelenti, hogy a felhasználók az MI-re bízzák a döntéshozatalt, anélkül hogy felülvizsgálnák az eredményeket. Ez a jelenség komoly következményekkel járhat, mivel a hibás vagy elfogult algoritmusok téves döntéseket eredményezhetnek, ami további társadalmi és gazdasági egyenlőtlenséghez vezethet. A Stanford Egyetem kutatása szerint az MI-algoritmusok által adott magyarázatok használata csökkentheti a túlzott mértékű támaszkodást az MI-rendszerekre, különösen akkor, ha ezek a magyarázatok egyszerűbbek és érthetőbbek, mint maga a feladat (Vasconcelos et al., 2022). A Microsoft Research jelentése szerint a felhasználók gyakran túlzott mértékben támaszkodnak az MI-rendszerekre, még akkor is, ha az így szerzett információk nem mindig pontosak vagy megbízhatók (Microsoft Research, 2022). A Brookings Intézet rámutatott, hogy az egészségügyben is jelentős kockázatot jelenthet az MI túlzott használata, mivel a hibás MI-rendszerek sérülésekhez és egyéb egészségügyi problémákhoz vezethetnek (Brookings, 2020). A Harvard Kennedy School kiemeli az adatvédelmi és biztonsági kockázatokat, valamint az MI-modellekben rejlő elfogultságokat, amelyek a politikai döntéshozatalban is megjelenhetnek (Harvard Kennedy School, 2024). Mint említettük, az MI-rendszerek használatának etikailag helyes módja az, hogy döntéstámogató eszközként alkalmazzák őket, nem pedig kizárólagos döntéshozóként. Az MI-alapú technológiák igazságos és nem diszkriminatív fejlesztése, valamint az emberi felülvizsgálat és beavatkozás segíthet csökkenteni az elfogultságot, és növelheti a döntéshozatali folyamatok megbízhatóságát.

Mechanikai megerősítést is létre kell hozni az MI-alapú technológiák szabályozására és az előre látható elfogultság és diszkrimináció megelőzésére. Ez azt jelenti, hogy olyan technikai megoldásokra van szükség, amelyek beépülnek az MI-rendszerekbe, hogy azok igazságosan és pártatlanul működjenek. A mechanikai megoldások mellett etikai alkalmazásokat is fel kell használni az MI javítására, különösen az értékeire és elveire vonatkozóan (Hagendorff, 2020). Azonban nagy hangsúlyt kell fektetni a technológiai részletekre, biztosítva, hogy az elvont etikai elvek technikailag is megfelelően legyenek beépítve az MI-rendszerekbe. Keng Siau és Weiyu Wang (2020) kiemelik, hogy az autonómiához, a szabadsághoz és az egyenlőséghez való jogot tiszteletben kell tartani az MI-alapú technológiák használata során. Az MI-rendszerek megfelelő működésének biztosítása érdekében szükség van az elszámoltathatóságra és felelősségre. Ennek érdekében létre kell hozni olyan eszközöket, amelyek megfelelnek az emberekre vonatkozó etikai normáknak és elveknek (Li et al., 2022). Azonban e cél nem érhető el, ha az eszköz fejlesztése során elfogultság van jelen. Ezért az MI fejlesztőinek etikai kötelességük olyan eszközöket alkotni, amelyek megfelelnek az etikai normáknak, és elkerülik az embereknek okozott előre nem látott károkat (Stahl, 2021). Ehhez univerzális etikai normák és az MI etikai alapelveinek megértése szükséges.

MUNKAVÁLLALÓK ÁTHELYEZÉSE ÉS GAZDASÁGI EGYENLŐTLENSÉG

Az MI-alapú technológiák jelentős változásokat hoztak a foglalkoztatási szektorban, ami munkahelyek kiszorításához vagy munkavállalók áthelyezéséhez vezetett. Az MI-alapú technológiák automatizált rendszereket használnak olyan feladatok elvégzésére, amelyeket korábban emberek láttak el. Ezeket a rendszereket gyors, hatékony és áramvonalas teljesítménnyel társítják, ami arra ösztönzi a szervezeteket, hogy az emberi munkaerő helyett fejlettebb és korszerűbb technológiákba fektessenek (Mu, 2023). Mint láttuk, az MI-eszközök széles körű adatfelhasználása hatalmas és komplex adatmennyiség elemzését teszi lehetővé, amely technikai alapot nyújt a döntéshozatalhoz szükséges mélyebb betekintésekhez. Thilo Hagendorff (2020) kiemeli, hogy az MI-rendszerek elemző kapacitása nem csupán etikai, hanem alapvetően technikai szempontból is lényeges, hiszen ezek az eszközök olyan komplex összefüggéseket tárnak fel, amelyek korábban, hagyományos módszerekkel nem voltak elérhetőek.

Az MI egyik fő alkalmazási területe a közlekedési szektor, ahol önvezető járműveket fejlesztettek ki. Idetartozik az Amerikai Egyesült Államokban folyamatos fejlesztés és kísérletezés alatt álló vezető nélküli autók (Loke, 2019). Ebben a termelési szektor is jelentős mértékben érintett. Nurhasyimah Abd Aziz és munkatársai (2021) beszámolója szerint az MI-alapú technológiákat aktívan használják

Az MI-eszközök széles körű adatfelhasználása hatalmas és komplex adatmennyiség elemzését teszi lehetővé, amely technikai alapot nyújt a döntéshozatalhoz szükséges mélyebb betekintésekhez.

a minőség-ellenőrzésben, a terméktervezés optimalizálásában és a folyamatfejlesztésben, ami szignifikáns mértékben csökkentette az emberi munkaerő szükségességét.

Az MI-alapú technológiák a tudásintenzív munkákra is hatással vannak, elég csak arra gondolnunk, hogy egyes MI-alapú rendszerek képesek kiváltani a pénzügyi és egészségügyi szektorban végzett tudásigényes munkákat. Például az MI-alapú pénzügyi tanácsadók (*robo-advisors*) képesek automatizált befektetési tanácsadást nyújtani (Lam, 2016). Ezek a rendszerek nagy mennyiségű pénzügyi adatot elemeznek, és algoritmusok segítségével személyre szabott befektetési portfóliókat javasolnak. Sőt, az MI-alapú diagnosztikai eszközök, mint például az IBM Watson for Oncology, képesek nagyszámú orvosi adatot elemezni, és ennek nyomán javaslatot tenni a rák kezelésére (IBM, 2023). Ezek a rendszerek az orvosi szakirodalomból és a betegadatokból származó információk alapján tesznek ajánlásokat, segítve az onkológusokat a kezelési döntésekben, csökkentve az orvosok terhelését, és növelve a diagnosztikai pontosságot.

Az MI hozzájárult rengeteg munkahely elvesztéséhez, mivel a repetitív munkákat folyamatosan automatizálják, különösen a gyártóiparban, az egészségügyi és pénzügyi szektorban, valamint az ügyfélszolgálati ágazatban. Masriadi és munkatársai (2023) szerint a munkavállalók áthelyezése szakaszokban történik. Az első szakaszban a munkálatók a monoton feladatokat bízzák az MI mechanikai rendszereire, hogy csökkentsék a működési költségeket. A második szakaszban az MI az analitikai munkákban kezdi felváltani az embereket, beleértve az adatelemzést és a döntéshozatalt is. Ugyanis csak attól, hogy egy munkavállaló „jóváhagy” egy MI-alapú technológia által felkínált döntést, az ilyenfajta alkalmazás nem nevezhető valódi döntéstámogatásnak. Amennyiben igaz, hogy az MI-alapú technológiák könnyen helyettesíthetik az embereket az adatelemzés és az üzleti analitika feladataiban (Zirar et al., 2023), akkor olyan intézkedések, mint a kötelező munkahelyi tréningek vagy egyéb átképzési programok segíthetnek minimalizálni az áthelyezések és munkahelyvesztések negatív következményeit. A harmadik szinten az MI-alapú technológia az embereket nemcsak az előző két területen váltja fel, hanem az emberekkel való adaptív interakciókban is helyettesítőként jelenik meg. A negyedik szakaszban az MI-eszközöket érzelmi intelligenciát igénylő feladatokhoz használják (Masriadi et al., 2023). Az ötödik és végső szakaszban az MI-alapú technológia olyan fejletté válik, hogy már képes az összes fent említett feladat elvégzésére, miközben a saját intellektuális képességeit fejleszti. Ezen a ponton pedig egyes területeken várhatóan teljesen felváltja a munkavállalót.

Az MI kiváltotta munkahelyi áthelyezések számos problémát vetettek fel a társadalmi-gazdasági hatások, az igazságosság és egyenlőség, valamint az emberi életre gyakorolt általános hatások vonatkozásában. Az áthelyezések negatívan befolyásolják a munkavállalók jóllétét, mivel hatással vannak az önbecsülésükre, a gazdasági teljesítményükre és a szükségleteik kielégítésére való képességükre. Ugyanakkor az MI-alapú technológiák által kiváltott munkaerő-áthelyezések olyan új foglalkoztatási lehetőségek megteremtéséhez vezethetnek, amelyek gazdasági szempontból inkább pozitív katalizátorként

hathatnak. Egyenlőtlenséghez vezethet az MI által létrehozott készséghiány, amely azt jelenti, hogy sok ember nem rendelkezik az MI-rendszerek kezeléséhez szükséges készségekkel, továbbá sokaknak nincs lehetőségük a megfelelő képzésre. Akik nem tudják kezelni az MI-t, hátrányba kerülnek a munkaerőpiacon, és ez jövedelmi különbségeket eredményez. Az MI alkalmazása a toborzási folyamatokban is gazdasági egyenlőtlenséghez vezethet. A gyorsan fejlődő MI-alapú toborzási rendszerek előítéletesek lehetnek a kisebbségi közösségekkel és az alacsonyabb társadalmi osztályokkal szemben, ami megnehezíti számukra a tisztességes foglalkoztatást. Nicholas Tilmes (2022) szerint az MI toborzási rendszereinek gyors fejlődése ahelyett, hogy átalakítaná a foglalkoztatási szektort, növeli a kisebbségi közösségekkel és az alacsonyabb osztályokkal szemben megnyilvánuló diszkriminációt. Az MI toborzási elfogultsága hozzájárul a gazdasági egyenlőtlenségekhez, mivel korlátozza a tisztességes foglalkoztatási kezdeményezéseket.

Erre a problémára az egyik legjobb megoldás lehet, ha újragondolják az MI-alapú technológia alkalmazását. Néhány területen e technológia használata szükséges lehet, de nem biztos, hogy valóban ez a legmegfelelőbb módja egy adott probléma orvoslásának, ugyanis bizonyos kihívásokat, amelyek az MI-alapú technológiákkal is kezelhetők, feltételezhetően jobban meg lehet oldani nem technológiai, azaz emberi beavatkozásokkal is. Példa erre a pszichológiai tanácsadás, illetve a mentális egészség területe. Bár léteznek MI-alapú chatbotok és virtuális asszisztensek, amelyek képesek alapvető tanácsokat nyújtani és támogatást biztosítani a felhasználóknak, az emberi pszichológusok és terapeuták empatikus és személyre szabott megközelítése sokkal hatékonyabb lehet a mentális egészség problémáinak kezelésében.

Összességében elmondható, hogy egyre növekvő szükség van annak meghatározására, hogy melyek a szükségtelen vagy legalábbis kevésbé szükséges MI-alkalmazási területek. Ennek pozitív hozadéka lenne, hogy a beazonosított területeken lényegesen visszaesne a munkahelyek kiszorítása és a munkavállalók áthelyezésével kapcsolatos kihívások száma, miközben az MI fejlesztői a gyümölcsözőbb alkalmazási területek felé fordíthatnák energiájukat (Prem, 2024).

Az MI-technológiák bevezetése jelentős gazdasági egyenlőtlenségeket okoz a munkaerőpiac átalakításával: automatizálja a rutinszerű és komplex feladatokat, ami sok munkahely megszűnéséhez és átalakulásához vezet, új készségek elsajátítására kényszerítve a munkavállalókat. Azok, akik nem tudnak időben alkalmazkodni, gyakran elveszítik állásukat, ami jövedelmi egyenlőtlenségekhez vezet. Ezzel szemben az MI-technológiák előnyeit főként a magas képzettségű, jól fizetett pozíciókban dolgozók élvezik, akik rendelkeznek a szükséges készségekkel és erőforrásokkal az MI-rendszerek fejlesztéséhez és használatához.

Mindemellert az MI-rendszerek bevezetéséhez jelentős tőke és infrastruktúra szükséges, amelyet főként nagyvállalatok és jól finanszírozott szervezetek tudnak biztosítani. Így a technológiai fejlődés hasznát leginkább ezek a nagyobb és gazdagabb vállalatok aratják le, míg a kisebb cégek és az alacsonyabb képzettségű munkavállalók hátrányos

helyzetbe kerülnek (Acemoglu–Restrepo, 2018). Erre példa, hogy a nemzetközi ügyvédi irodák jelentős előnyt élveznek az egyéni ügyvédekkel és kisebb irodákkal szemben az MI-eszközök kihasználásában, nagyrészt a nagyobb pénzügyi forrásaiknak és infra-

Az MI-alapú technológiák előnyeit elsősorban azok élvezik, akik rendelkeznek a megfelelő készségekkel, tőkével és tudással ezek kiaknázásához.

struktúrájuknak köszönhetően. Ezek a nagy irodák megengedhetik maguknak, hogy fejlett, saját MI-rendszerekbe fektessenek, amelyek számos feladatot automatizálnak, például a jogi kutatást vagy a dokumentum-ellenőrzést. Annak képessége, hogy MI-t integráljanak a működésükbe, nemcsak a szolgáltatás minőségét javítja, hanem jelentősen csökkenti az átfutási időt is, versenyelőnyt biztosítva a nagyobb irodáknak (Chen et al., 2022;

Dennehy et al., 2023). Ezzel szemben a kisebb irodák és az egyéni ügyvédek gyakran pénzügyi és logisztikai korlátokkal szembesülnek, amelyek lehetetlenné teszik hozzáférésüket ezekhez a fejlett technológiákhoz. Bár az MI kiegyenlíthetné a versenyfeltételeket azáltal, hogy javítja a hatékonyságot, az ezekhez az eszközökhöz szükséges kezdeti befektetés és folyamatos költségek megterhelők lehetnek.

Azok a személyek és szervezetek, akik/amelyek irányítják az MI-rendszereket, vagy hozzáférésük van ezekhez, jelentős versenyelőnyre tesznek szert (Lu–Zhou, 2021). Az automatizálás révén csökkenthetik működési költségeiket, növelhetik termelékenységüket, és javíthatják szolgáltatásaik minőségét. Ezzel szemben azok, akik nem tudják alkalmazni ezeket a technológiákat, lemaradnak a versenyben, és képtelenek lesznek ugyanolyan hatékonysággal és minőségben működni. Ennek következtében a jövedelmi és gazdasági szakadék tovább növekszik a technológiahasználók és a technológiahasználaton kívül rekedtek között.

Pozitívumként említendő, hogy az MI alkalmazása jelentős hatással van a termelékenységre és az innovációra, ami hozzájárulhat a gazdasági növekedéshez. Ugyanakkor ez a növekedés nem egyenletesen oszlik meg, mivel a technológia alkalmazása különböző szektorokban eltérő mértékben történik. Mint már említettük, az MI-alapú technológiák előnyeit elsősorban azok élvezik, akik rendelkeznek a megfelelő készségekkel, tőkével és tudással ezek kiaknázásához. Ez a folyamat korábban is megfigyelhető volt a technológiai fejlődés során, de az MI esetében még hangsúlyosabbá vált (Heilinger, 2022).

KONKLÚZIÓ

Az MI-technológiák fejlődése és alkalmazása számos etikai dilemmát vet fel, amelyek az alapos megfontolás mellett megfelelő szabályozási keretek kialakítását igénylik. Az MI-rendszerek gyors térhódítása az élet különböző területein – az egészségügytől a közlekedésen át a pénzügyi szektorig – nemcsak technológiai előnyöket hozott magával, hanem komoly társadalmi és gazdasági kihívásokat is. Az előítéletesség, diszkrimináció,

a munkavállalók áthelyezése és a gazdasági egyenlőtlenségek olyan etikai dilemmákat vetnek fel, amelyek már az MI-technológiák megjelenése előtt is léteztek, de azok alkalmazása fokozhatja e problémákat. A tanulmány azoknak az etikai kérdéseknek a vizsgálatára fókuszált, amelyek az MI-rendszerek alkalmazásával új kihívásokat állítanak elénk.

Az előítéletesség és a diszkrimináció kérdéskörével kapcsolatban világossá vált, hogy az MI-rendszerek gyakran továbbviszik a tanító adatokban meglévő torzításokat, ami igazságtalansághoz vezethet a különböző társadalmi csoportok között. Az olyan példák, mint az Amazon cég toborzási rendszere, a PredPol bűnügyi előre jelző rendszer és a COMPAS kockázatbecslő rendszer, mind rámutatnak, hogy az MI-algoritmusok hibás adatokkal és előítéletes programozással súlyos társadalmi problémákat generálhatnak.

A munkavállalók áthelyezése és a gazdasági egyenlőtlenségek vonatkozásában kiemelt, hogy az MI-technológiáknak köszönhető automatizálás jelentős munkahelyvesztést és jövedelmi különbségeket eredményez. Az MI-rendszerek bevezetése főként a nagyvállalatok és a magasan képzett szakemberek számára hoz előnyöket, míg a kisebb cégek és az alacsonyabb képzettségű munkavállalók hátrányba kerülnek. Az MI alkalmazásának újragondolása és a nem technológiai megoldások előnyben részesítése bizonyos területeken segíthet abban, hogy csökkentsük ezeket az egyenlőtlenségeket és a munkahelyi áthelyezések negatív hatásait.

Összességében az MI-technológiák megjelenése által felvetett etikai kérdések kezelése érdekében elengedhetetlen a szabályozási keretek kidolgozása és az etikai normák szigorú betartása. Erre lehetőséget ad a kockázatalapú, keretrendszer-jellegű szabályozási modell. Az MI-rendszerek fejlesztése és alkalmazása során biztosítani kell, hogy ezek igazságosak, átláthatók és elszámoltathatók legyenek. Emellett szükség van az MI-alapú technológiák előnyeinek egyenlő elosztására és a képzési lehetőségek biztosítására, hogy mindenki számára hozzáférhetővé váljanak az új technológiák és előnyeik. Csak így érhetjük el, hogy az MI valóban a társadalom egészének javát szolgálja, és ne mélyítse tovább a meglévő társadalmi és gazdasági egyenlőtlenségeket.

Felhasznált irodalom

- ABD AZIZ, NURHASYIMAH – ADNAN, NUR A. A. – ABD WAHAB, DZURAI DAH – AZMAN, ABDUL HADI (2021): Component design optimization based on artificial intelligence in support of additive manufacturing repair and restoration: Current status and future outlook for remanufacturing, *Journal of Cleaner Production*, 296, 126401. DOI: <https://doi.org/10.1016/j.jclepro.2021.126401>
- ACEMOGLU, DARON – RESTREPO, PASCUAL (2018): Artificial Intelligence, Automation, and Work, *NBER Working Paper*, 24196. DOI: <https://doi.org/10.3386/w24196>
- ANGWIN, JULIA – LARSON, JEFF – MATTU, SURYA – KIRCHNER, LAUREN (2016): Machine bias, *ProPublica*, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (letöltve: 2024. 04. 25.).
- BERK, RICHARD A. (2021): Artificial intelligence, predictive policing, and risk assessment for law enforcement, *Annual Review of Criminology*, 4, 209–237. DOI: <https://doi.org/10.1146/annurev-criminol-051520-012342>
- BROOKINGS (2020): Risks and remedies for artificial intelligence in health care, <https://www.brookings.edu/research/risks-and-remedies-for-artificial-intelligence-in-health-care/> (letöltve: 2024. 07. 03.).
- BYRNE, MATTHEW D. (2021): Reducing bias in healthcare artificial intelligence, *Journal of PeriAnesthesia Nursing*, 36 (3), 313–316. DOI: <https://doi.org/10.1016/j.jpnan.2021.03.009>
- CHEN, DONGHUA – ESPERANÇA, JOSÉ PAULO – WANG, SHAO FENG (2022): The impact of artificial intelligence on firm performance: An application of the resource-based view to e-commerce firms, *Frontiers in Psychology*, 13. DOI: <https://doi.org/10.3389/fpsyg.2022.884830>
- CHEN, ZHISHENG (2023): Ethics and discrimination in artificial intelligence-enabled recruitment practices, *Humanities and Social Sciences Communications*, 10 (1), 1–12. DOI: <https://doi.org/10.1057/s41599-023-02079-x>
- COECKELBERGH, MARK (2020): *AI Ethics*, Cambridge, MIT Press. DOI: <https://doi.org/10.7551/mitpress/12549.001.0001>
- DASTIN, JEFFREY (2018): Amazon scraps secret AI recruiting tool that showed bias against women, *Reuters*, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> (letöltve: 2024. 04. 25.).
- DENNEHY, DENIS – GRIVA, ANASTASIA – POULOU DI, NANCY – DWIVEDI, YOGESH K. – MÄNTYMÄKI, MATTI – PAPPAS, ILIAS O. (2023): Artificial intelligence capability and firm performance: A sustainable development perspective by the mediating role of data-driven culture, *Information Systems Frontiers*. DOI: <https://doi.org/10.1007/s10796-022-10365-3>
- DRESSEL, JULIA – FARID, HANY (2018): The accuracy, fairness, and limits of predicting recidivism, *Science Advances*, 4 (1), eaao5580. DOI: <https://doi.org/10.1126/sciadv.aao5580>

- ELMORE, MATTHEW (2023): The hidden costs of ChatGPT: A call for greater transparency, *The American Journal of Bioethics*, 23 (10), 47–49. DOI: <https://doi.org/10.1080/15265161.2023.2250335>
- EUR-LEX (2021): Javaslat az Európai Parlament és a Tanács Rendelete a Mesterséges Intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról, COM/2021/206 final, <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX%3A52021PC0206> (letöltve: 2024. 05. 10.).
- FERRER, XAVIER – VAN NUENEN, TOM – SUCH, JOSE M. – COTÉ, MARK – CRIADO, NATALIA (2021): Bias and discrimination in AI: a cross-disciplinary perspective, *IEEE Technology and Society Magazine*, 40 (2), 72–80. DOI: <https://doi.org/10.1109/MTS.2021.3056293>
- HAGENDORFF, THILO (2020): The ethics of AI ethics: An evaluation of guidelines, *Minds and Machines*, 30 (1), 99–120. DOI: <https://doi.org/10.1007/s11023-020-09517-8>
- HALE, CRAIG (2023): ChatGPT is having a really bad impact on the environment, <https://www.techradar.com/pro/chatgpt-is-having-a-really-bad-impact-on-the-environment> (letöltve: 2024. 04. 28.).
- HARVARD KENNEDY SCHOOL (2024): Assessing Risks and Best Practices of AI Large Language Models in Policymaking, <https://www.hks.harvard.edu/events/assessing-risks-and-best-practices-ai-large-language-models-policymaking> (letöltve: 2024. 07. 03.).
- HEILINGER, JAN-CHRISTOPH (2022): The ethics of AI ethics. A constructive critique, *Philosophy & Technology*, 35 (3), 61. DOI: <https://doi.org/10.1007/s13347-022-00557-9>
- IBM (2023): 5725-W51 IBM Watson for Oncology, <https://www.ibm.com/docs/en/announcements/watson-oncology?region=CAN> (letöltve: 2024. 05. 01.).
- LAM, JONATHAN WALTER (2016): *Robo-Advisors, A Portfolio Management Perspective*, New Haven, Yale College, https://economics.yale.edu/sites/default/files/2023-01/Jonathan_Lam_Senior%20Essay%20Revised.pdf (letöltve: 2024. 04. 15.).
- LI, FAN – RUIJS, NICK – LU, YUAN (2022): Ethics & AI: A systematic review on ethical concerns and related strategies for designing with AI in healthcare, *AI*, 4 (1), 28–53. DOI: <https://doi.org/10.3390/ai4010003>
- LOKE, SENG W. (2019): Cooperative automated vehicles: A review of opportunities and challenges in socially intelligent vehicles beyond networking, *IEEE Transactions on Intelligent Vehicles*, 4 (4), 509–518. DOI: <https://doi.org/10.1109/TIV.2019.2938107>
- LU, YINGYING – ZHOU, YIXIAO (2021): A review of the economics of artificial intelligence, *Journal of Economic Surveys*, 35 (4), 1045–1072. DOI: <https://doi.org/10.1111/joes.12422>

- LUM, KRISTIAN – ISAAC, WILLIAM (2016): To predict and serve?, *Significance*, 13 (5), 14–19. DOI: <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- MASRIADI – DASMADI – EKANINGRUM, NURNI EFRI – HIDAYAT, MUHAMMAD SYAHRUL – YULIATY, FARIDA (2023): Exploring the future of work, Impact of automation and artificial intelligence on employment, *Endless: International Journal of Future Studies*, 6 (1), 125–136. DOI: <https://doi.org/10.54783/etj.v6i1.131>
- MICROSOFT RESEARCH (2022): Overreliance on AI, <https://www.microsoft.com/en-us/research/uploads/prod/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf> (letöltve: 2024. 07. 03.).
- MU, WEIJIA (2023): How artificial intelligence affects workforces: The impact of biased recruitment and job displacement risk, *Highlights in Business, Economics and Management*, 23, 19–25. DOI: <https://doi.org/10.54097/2t4h0q42>
- PREM, ERICH (2024): Approaches to ethical AI, in WERTHNER, HANNES ET AL. (szerk.): *Introduction to Digital Humanism*, Cham, Springer. DOI: https://doi.org/10.1007/978-3-031-45304-5_15
- SIAU, KENG – WANG, WEIYU (2020): Artificial intelligence (AI) ethics: ethics of AI and ethical AI, *Journal of Database Management*, 31 (2), 74 Springer, 87. DOI: <https://doi.org/10.4018/JDM.2020040105>
- STAHL, BERND CARSTEN (2021): Addressing Ethical Issues in AI, in STAHL, BERND CARSTEN: *Artificial Intelligence for a Better Future*, Cham, Springer. DOI: https://doi.org/10.1007/978-3-030-69978-9_5
- STATE COUNCIL OF PRC (2017): China issues guideline on Artificial Intelligence development, https://english.www.gov.cn/policies/latest_releases/2017/07/20/content_281475742458322.htm (letöltve: 2024. 05. 10.).
- TILMES, NICHOLAS (2022): Disability, fairness, and algorithmic bias in AI recruitment, *Ethics and Information Technology*, 24 (21). DOI: <https://doi.org/10.1007/s10676-022-09633-2>
- VASCONCELOS, HELENA – JÖRKE, MATTHEW – GRUNDE-McLAUGHLIN, MADELEINE – GERSTENBERG, TOBIAS – BERNSTEIN, MICHAEL S. – KRISHNA, RANJAY (2022): Explanations can reduce overreliance on AI systems during decision-making, *Proceedings of the ACM on Human-Computer Interaction*. DOI: <https://doi.org/10.48550/arXiv.2212.06823>
- WHITE HOUSE (2023): FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/> (letöltve: 2024. 05. 10.).
- ZIRAR, ARAZ – ALI, SYED IMRAN – ISLAM, NAZRUL (2023): Worker and workplace Artificial Intelligence (AI) coexistence: Emerging themes and research agenda, *Technovation*, 124, 102747. DOI: <https://doi.org/10.1016/j.technovation.2023.102747>

Bias, Discrimination, and Economic Inequalities

Ethical Challenges of Artificial Intelligence

ABSTRACT

The rapid development and proliferation of AI technologies raise numerous ethical questions that require thorough examination and regulation. This study aims to map and discuss the key social and ethical concerns related to AI applications. It divides the ethical dilemmas of AI technologies into two main areas: bias and discrimination, and job displacement and economic inequalities. AI systems often perpetuate biases present in training data, leading to injustices among various societal groups. Additionally, AI automation results in significant job loss and income inequalities, as the technological benefits are primarily enjoyed by large corporations and highly skilled professionals. The study presents several examples of ethical challenges in AI applications and offers recommendations for the development of regulatory frameworks and ethical standards.

KEYWORDS: artificial intelligence, ethical dilemmas, bias, discrimination, economic inequalities

MIKE NIMRÓD | A beépített adatvédelem és a mesterséges intelligencia szabályozási kérdéseinek szakértője, jelentős kutatási tapasztalattal rendelkezik a technológia társadalmi és gazdasági hatásainak területén. Doktori fokozatát a Budapesti Corvinus Egyetemen szerezte, számos publikációja és előadása foglalkozik a személyes adatok védelme, valamint az alkalmazott MI szabályozási és etikai kihívásaival.