KOMMUNIKÁCIÓ, KÖZVÉLEMÉNY, MÉDIA

2024/4. szám

AZ ÁLHÍREK KORA: MESTERSÉGES INTELLIGENCIA ÉS ETIKA A DIGITÁLIS TÉRBEN

Fenyvesi Judit
jfenyvesi@gmail.com

DOI: 10.20520/JEL-KEP.2024.4.43

Absztrakt

A digitális korszakban az álhírek terjedése és a mesterséges intelligencia (MI) alkalmazása számos etikai kihívást vet fel. Az információáramlás felgyorsulásával és a közösségi média dominanciájával az álhírek könnyebben terjednek és nagyobb hatást érhetnek el, ami komoly fenyegetést jelent a nyilvános diskurzus hitelességére. Az MI mindkét irányba hat: egyrészt elősegítheti a dezinformáció terjedését, másrészt segíthet annak felismerésében és visszaszorításában. Az álhírek kérdéskörének elemzése így a kommunikációs technológiák fejlődésével összefüggő etikai dilemmákra is rámutat. A tanulmány az MI és az álhírek kapcsolatát három fő szempontból tárgyalja: az álhírek terjesztését segítő MI technológiák, az álhírek felismerésére használt MI-alapú megoldások, valamint az ezekkel összefüggő etikai kérdések.

Kulcsszavak

álhírek, mesterséges intelligencia, etika, digitális tér, közösségi média

THE AGE OF FAKE NEWS: ARTIFICIAL INTELLIGENCE AND ETHICS IN THE DIGITAL SPHERE

Fenyvesi Judit

Abstract

In the digital age, spreading fake news and artificial intelligence (AI) applications raises numerous ethical challenges. With the acceleration of information flow and the dominance of social media, fake news spreads more easily and can have a greater impact, posing a significant threat to the credibility of public discourse. AI influences this issue in both directions: on the one hand, it can facilitate the dissemination of disinformation, while on the other hand, it can assist in its detection and containment. Analyzing the issue of fake news thus highlights the ethical dilemmas associated with the development of communication technologies. This study examines the relationship between AI and fake news from three main perspectives: AI technologies that aid the dissemination of fake news, AI-based solutions for detecting fake news, and the ethical questions related to these issues.

Keywords

fake news, artificial intelligence, ethics, digital space, social media

JEL-KÉP

KOMMUNIKÁCIÓ, KÖZVÉLEMÉNY, MÉDIA

2024/4. szám

AZ ÁLHÍREK KORA: MESTERSÉGES INTELLIGENCIA ÉS ETIKA A DIGITÁLIS TÉRBEN

Fenyvesi Judit

jfenyvesi@gmail.com

Bevezetés

A digitális média térnyerése jelentős mértékben átalakította az információáramlás szerkezetét és dinamikáját. A hagyományos médiacsatornák által közvetített, viszonylag szűrt információk helyét egy decentralizált, gyorsan terjedő és gyakran kontrollálhatatlan hírfolyam vette át. Ez az átalakulás egyszerre nyitott új lehetőségeket az információkhoz való hozzáférésben, másrészt az álhírek által okozott problémák egyre súlyosabb kihívásként jelentkeznek, amelyek alapvetően érintik a demokratikus társadalmak információs ökoszisztémáját is. Az álhírek megjelenése és elterjedése különösen problematikus a közvélemény formálásában, mivel ezek az információk nemcsak a társadalmi bizalmat rombolják, hanem az egyéni döntéshozatalt is negatívan befolyásolják (Aczél 2017). A közösségi média térhódításával az álhírek nemcsak gyorsabban terjednek, hanem személyre szabott algoritmusok révén célzottan érnek el különböző közönségeket.

Az álhírek a modern információs társadalom egyik legsúlyosabb problémájává váltak. A hazai és a nemzetközi szakirodalomban az álhírek (fake news) fogalmára számtalan meghatározás található, ami félreértéseket is okozhat. Wardle és Derakhshan (2017) az „információs zavar” fogalmát vezették be, amely lehetővé teszi az álhírek szélesebb körű értelmezését. Három fő típust különböztetnek meg: 1. a félretájékoztatást (misinformation), a hamis vagy félrevezető információk megosztását, amely mögött nem áll szándékos károkozási szándék; 2. a dezinformációt (disinformation), a szándékosan félrevezető, hamis információk terjesztését károkozási céllal, és 3. a rosszindulatú információt (malinformation), a valós információk károkozási céllal történő manipulálását és megosztását. Ezek a kategóriák segítenek pontosítani az álhírek természetét, és meghatározni, hogyan illeszkednek az információs ökoszisztémába. Fehér Katalin (2017) szerint az álhírek nemcsak a hírterjesztés hagyományos logikáját írják felül, hanem az igazság fogalmának értelmezését is megkérdőjelezik.

Az álhírek terjedésében a mesterséges intelligencia (MI) kiemelt szerepet játszik. Az MI olyan technológiák összessége, amelyek képesek nagy mennyiségű adat elemzésére, mintázatok felismerésére és akár önálló döntéshozatalra (Floridi 2023). A mesterséges intelligencia arra is alkalmas, hogy gyorsan állítson elő és terjesszen hitelesnek látszó, de hamis információkat, ugyanakkor hatékony eszköz lehet az álhírek azonosításában és visszaszorításában. Fehér Katalin (2018) úgy véli, hogy az MI alapvetően átalakítja a digitális valóságot, nem csupán passzív információfeldolgozó, hanem aktív narratívateremtő, amely az adatokat történetekké alakítja. Ma már elsődleges hírforrásunk, a közösségi média algoritmusai

szintén az MI alkalmazására épülnek, és gyakran a felhasználói elköteleződés maximalizálására optimalizáltak. Ez azonban gyakran az érzelmileg töltött, polarizáló tartalmak előtérbe helyezéséhez vezet, amelyek magukban foglalhatják az álhíreket is (Allcott – Gentzkow 2017).

A mesterséges intelligencia alkalmazása számos etikai kérdést is felvet. Az MI-etikai kérdések középpontjában az áll, hogyan alkalmazhatók ezek a rendszerek felelősségteljesen, különös tekintettel az álhírek és dezinformációk felismerésére. Az egyik legfontosabb probléma, hogy a mesterséges intelligencia döntéshozatali folyamatai nem mindig átláthatók, az MI-alapú döntéshozatali rendszerek gyakran úgynevezett „fekete dobozokként" működnek, amelyek belső logikája nem hozzáférhető sem a felhasználók, sem a szabályozók számára (Floridi 2023). Ez pedig komoly aggályokat vet fel a felelősség és az etikus működés terén. Az MI algoritmusok által hozott döntések továbbá hatással lehetnek a szólásszabadságra is, különösen akkor, ha a dezinformáció elleni küzdelem érdekében tartalmakat szűrnek vagy cenzúráznak. Zuboff (2019) figyelmeztet arra, hogy a technológiai cégek beavatkozása a tartalomkezelésbe hosszú távon korlátozhatja a demokratikus nyilvánosságot. Az MI alkalmazása az álhírek elleni küzdelemben ugyanakkor jelentős előnyöket is kínál, például az álhírek automatizált felismerése és szűrése terén. Mindemellett e technológiák társadalmi hatásait csak akkor lehet hatékonyan kezelni, ha a technológiai innovációt megfelelő etikai és szabályozási keretek között valósítják meg (O'Neil 2016).

Az álhírek terjesztését segítő MI technológiák

A mesterséges intelligencia által támogatott technológiák, mint a chatbotok, a deepfake videók, a mesterséges influencerek és a szöveggeneráló algoritmusok, alapvetően változtatják meg az információáramlás mechanizmusait és számos új lehetőséget biztosítanak az álhírek előállítására és terjesztésére. Ezek az eszközök nemcsak hatékonyabbá teszik a hamis információk előállítását és terjesztését, hanem kihívást jelentenek a hitelesség és a demokratikus diskurzus fenntartása szempontjából is. A közösségi média algoritmusai tovább fokozzák ezt a problémát, mivel előnyben részesítik az érzelmileg túlfűtött és polarizáló tartalmakat. Az alábbiakban bemutatom azokat az eszközöket és mechanizmusokat, amelyek különösen hatékonyak a dezinformáció terjesztésében.

Automatizált tartalomgyártás

A chatbotok fejlesztése és elterjedése a mesterséges intelligencia egyik rendkívül dinamikusan fejlődő területe, amelyek már nemcsak ügyfélszolgálatok vagy oktatási platformok részei, hanem a dezinformációs kampányok aktív eszközei is. Demeter Márton (2019) rámutat arra, hogy a chatbotok használata politikai és gazdasági manipulációk során egyre gyakoribbá válik, mivel képesek hiteles emberi interakciókat szimulálni. Ezek az algoritmusok különösen veszélyesek az információ hitelességére, mivel gyorsan és költséghatékonyan képesek személyre szabott üzenetek terjesztésére, hatékonyan formálhatják a közvéleményt. A 2024-es *Reuters Digital News Report* megjegyzi, hogy a chatbotok által generált álhírek különösen nagy hatást gyakorolnak a közösségi média fiatalabb felhasználóira, akik kevésbé szkeptikusak a tartalmakkal szemben. A chatbotok továbbá nemcsak információkat terjesztenek, hanem aktívan generálnak új narratívákat is, különösen politikai kampányok vagy társadalmi konfliktusok idején. Egyes kutatások szerint a chatbotok akár a közösségi média forgalmának jelentős részét is képesek manipulálni (Wardle – Derakhshan 2017).

A deepfake technológia a mesterséges intelligencia egyik igen fejlett és fenyegető eszköze. Aczél Petra (2023) hangsúlyozza, hogy a deepfake nemcsak vizuális manipulációt végez, hanem alapvetően megkérdőjelezi a valóság és a hitelesség fogalmát. Ez a technológia képes politikai szereplők lejáratására, bűncselekmények hamis bizonyítékainak létrehozására,

és társadalmi konfliktusok szítására. Andok Mónika (2022) szerint a deepfake technológia megváltoztatja a kommunikációs normákat, és olyan etikai kérdéseket vet fel, amelyek meghaladják a hagyományos médiakritikai kereteket. A társadalom nagy része nincs felkészülve arra, hogy felismerje az ilyen típusú manipulációkat, ami növeli ezek hatékonyságát. A deepfake-ek elleni védekezés egyre fontosabbá válik, azonban a védekezési algoritmusok fejlesztése gyakran lassabb, mint a támadó technológiák fejlődése. Shu et al. (2017) megjegyzik, hogy az MI-alapú védekezési rendszerek ugyan képesek bizonyos mértékig felismerni a deepfake-eket, de folyamatos innováció szükséges a technológiai versenyben maradáshoz. A deepfake technológia az MI vizuális manipulációs képességeit használja ki, hogy valósághű, de teljesen hamis videókat és képeket hozzon létre. Ez különösen veszélyes lehet, ha politikai vezetőkről, hírességekről vagy társadalmilag érzékeny helyzetekről készülnek deepfake tartalmak. Fehér Katalin (2018) is hangsúlyozza, hogy a deepfake technológia nemcsak az emberek valóságérzékelését torzítja, hanem alapvetően megkérdőjelezi az igazság fogalmát is. A deepfake-ek például felhasználhatók hamis vallomások, manipulált bizonyítékok előállítására vagy akár személyes lejáró kampányok indítására is.

A mesterséges intelligenciával létrehozott influencerek új jelenségként jelentek meg a közösségi média világában. Ezek a virtuális személyiségek, mint például Lil Miquela vagy Shudu Gram, népszerűek a fiatalabb generációk körében, és olyan hatást keltenek, mintha valódi emberek lennének. Aczél Petra (2023) szerint a mesterséges influencerek alkalmasak arra, hogy finoman és hitelesnek tűnő módon terjesszenek dezinformációt. Ezek az influencerek olyan kampányokban vesznek részt, amelyek a fogyasztói szokások, politikai nézetek vagy társadalmi kérdések manipulálására irányulnak.

Az MI egyik legnagyobb előnye, ugyanakkor veszélye is a szöveggeneráló algoritmusok fejlődése. Ezek az eszközök képesek összetett, koherens szövegek előállítására, amelyek gyakran nehezen különböztethetők meg az ember által írt tartalmaktól. Allcott és Gentzkow (2017) megállapították, hogy ezek az algoritmusok különösen hatékonyak a társadalmi polarizáció erősítésében, mivel érzelmileg túlfűtött és manipulatív tartalmakat generálnak. Brynjolfsson és McAfee (2014) szerint ezek az algoritmusok átalakítják az információs munkát, de ugyanakkor megkönnyítik a dezinformáció előállítását is. A GPT-alapú rendszerek például képesek személyre szabott dezinformációs kampányokat létrehozni, amelyek nemcsak gyorsabbak, de hatékonyabbak is a célzott közönség elérésében. Fehér Katalin (2018) megjegyzi, hogy az ilyen eszközök elterjedése alapvetően formálja át a digitális ökoszisztéma dinamikáját, különös tekintettel a hitelességre és az átláthatóságra. Fehér Katalin (2018) rámutat arra is, hogy a szöveggeneráló eszközök felhasználhatók a közösségi média befolyásolására, különösen a választási időszakokban.

Közösségi média algoritmusai és szűrőbuborékok

A közösségi médiaplatformok algoritmusai a felhasználók viselkedésére alapozott tartalmakat részesítik előnyben, amelyek gyakran érzelmileg túlfűtött és megosztó üzenetek (Allcott – Gentzkow 2017). Ez a jelenség erősíti a szűrőbuborékok (filter bubble) és visszhangkamrák (echo chambers) hatását (Pariser 2011). Pariser azt állítja, hogy ezek a buborékok jelentősen csökkentik a társadalmi diskurzus sokszínűségét, mivel kizárják az eltérő nézeteket. A felhasználók ugyanis hajlamosak azokat a híreket fogyasztani, amelyek megerősítik saját nézeteiket, miközben kizárják az ellentétes információkat. A visszhangkamrák tovább mélyítik ezt a problémát, mivel a hasonló nézeteket valló csoportok izolált információs környezetet hoznak létre. Andok Mónika (2015) szerint ezek a mechanizmusok különösen alkalmasak arra, hogy elősegítsék az álhírek széles körű terjedését, mivel a felhasználók ritkán kerülnek kapcsolatba az ellentétes álláspontokkal.

A közösségi média algoritmusai rejtetten működnek, és a felhasználók gyakran nem tudják, milyen paraméterek alapján kerülnek előtérbe az egyes tartalmak. Zuboff (2019) hangsúlyozza, hogy ez az átláthatatlanság komoly etikai dilemmákat vet fel, különösen akkor, ha az algoritmusok előnyben részesítik a polarizáló vagy félrevezető tartalmakat. A Reuters Institute Digital News Report arra is felhívja a figyelmet a fiatalabb generációk esetében, akik számára a közösségi média az elsődleges hírforrás, hogy az algoritmusok működése döntő szerepet játszik abban, milyen információkkal találkoznak. Ez tovább erősíti a platformok felelősségét az álhírek terjedésének megakadályozásában.

Geopolitikai manipuláció MI-technológiák segítségével

A mesterséges intelligencia technológiák geopolitikai konfliktusokban és nemzetközi hatalmi játszmákban való felhasználása egyre nagyobb aggodalomra ad okot. A dezinformációs kampányok, amelyeket állami szereplők vagy azok támogatói indítanak, gyakran MI-t használnak célzott narratívák terjesztésére. Wardle és Derakhshan (2017) kiemelik, hogy a mesterséges intelligenciával támogatott dezinformációs kampányokat az államok gyakran geopolitikai célokra használják, például választások befolyásolására vagy társadalmi zavargások szítására. Egy klasszikus példa az Internet Research Agency tevékenysége, amely orosz állami támogatással terjesztett álhíreket a 2016-os amerikai elnökválasztás során (DiResta et al. 2018). Fehér Katalin (2018) szerint a nemzetközi választások során alkalmazott MI-eszközök, például a szociális hálózatok célzott manipulációja, jelentősen torzíthatják a demokratikus döntéshozatalt. Az MI-alapú adatelemző rendszerek azonosíthatják a választók érzelmi hajlandóságait, és testreszabott üzeneteket hozhatnak létre, amelyek közvetlenül a választói magatartást célozzák. Zuboff (2019) hangsúlyozza, hogy a közösségi média algoritmusai gyakran előnyben részesítik az érzelmekre ható tartalmakat, amelyek tökéletes eszközök lehetnek geopolitikai propaganda terjesztésére is. Ezek az algoritmusok az érzelmi polarizációt kihasználva erősítik az állami vagy nemzetközi érdekeket szolgáló dezinformáció hatását. A Reuters Digital News Report (2024) kiemeli, hogy a geopolitikai álhírek terjedése különösen veszélyes, mivel a globális médiapiacon a helyi és nemzetközi szereplők közötti határok elmosódnak. Ez megnehezíti az információk hitelességének ellenőrzését és a források azonosítását.

Újonnan megjelenő technológiák az álhírek terjesztésében

A generatív audiómodellek, amelyek valósághű hangokat képesek létrehozni, újabb szintre emelik az álhírek hitelességét. Ezek az eszközök lehetővé teszik híresség, politikusok vagy akár hétköznapi emberek hangjának imitálását. Aczél Petra (2023) szerint ezek a technológiák különösen veszélyesek, mivel az emberek többsége nagyobb hitelességet tulajdonít a hallott információknak, mint a vizuális tartalmaknak. Ez a technológia például hamis telefonhívások vagy manipulált rádióinterjúk készítésére is felhasználható. Az MI prediktív elemző képességei pedig lehetővé teszik a célközönség viselkedési mintáinak előrejelzését. Ezáltal az álhírek terjesztése még célzottabbá és hatékonyabbá válik. Brynjolfsson és McAfee (2014) hangsúlyozzák, hogy ezek az eszközök nemcsak az aktuális tartalomgyártásra, hanem a hosszú távú dezinformációs stratégiák kialakítására is használhatók.

Az álhírek felismerése és visszaszorítása MI segítségével

Az álhírek felismerésére és visszaszorítására szolgáló MI-alapú technológiák, mint a természetes nyelvfeldolgozás, a tényellenőrzés, valamint a blokklánc-alapú hitelesítési rendszerek, effektív eszközök lehetnek a dezinformáció elleni küzdelemben. Az NLP

technológiák és a platformszintű megoldások hatékonyan azonosítják a dezinformációs mintázatokat és a hiteltelen forrásokat. Ugyanakkor ezeknek az eszközöknek a sikeres alkalmazásához szükséges a felhasználói tudatosság növelése, a társadalmi felelősségvállalás erősítése és a szabályozási keretek átláthatóbbá tétele.

Természetes nyelvfeldolgozás és MI-alapú algoritmusok

A természetes nyelvfeldolgozás (Natural Language Processing, NLP) az MI egyik legfontosabb területe, amely lehetővé teszi a gépek számára az emberi nyelv értelmezését, feldolgozását és elemzését. Az NLP technológiák különösen hatékonyak a dezinformációs mintázatok azonosításában, mivel képesek automatizáltan elemezni a szövegeket, azonosítva a félrevezető narratívákat és a hiteltelen forrásokat (Shu et al. 2017). A mesterséges intelligencia ezen alkalmazása különösen fontos lehet a közösségi média platformok számára, ahol a hamis információk gyorsan elérhetik a széles közönséget. Pennycook és Rand (2019) megállapította, hogy az NLP technológiák hatékonysága jelentősen növelhető, ha az adatbázisokban szereplő információkat folyamatosan frissítik, és a gépi tanulási modellek tanulási folyamatait rendszeresen finomítják, például egyes algoritmusok képesek azonosítani azokat a kulcsszavakat és szövegstruktúrákat, amelyek jellemzőek az álhírekre. A közösségi média platformok és a keresőmotorok szintén alkalmazzák az NLP-alapú algoritmusokat a tartalomszűrésben. Például a Facebook és a Google olyan rendszereket fejlesztett ki, amelyek képesek felismerni a dezinformációt, és figyelmeztetéseket jelenítenek meg a kétes forrásokra (Pennycook – Rand 2019). Az ilyen rendszerek azonban gyakran átláthatatlan módon működnek, ami megnehezíti a felhasználók számára, hogy megértsék, miért tekintik az adott tartalmat álhírnek (Floridi 2023).

Platformszintű megoldások

A közösségi média platformok, mint a Facebook, Twitter és YouTube, erőfeszítéseket tesznek az álhírek terjedésének visszaszorítása érdekében, például figyelmeztetéseket jelenítenek meg a kétes forrásoknál (Pennycook – Rand 2019). Ugyanakkor ezek az eszközök gyakran átláthatatlan módon működnek, ami bizalomvesztést eredményez (Floridi 2023). Az MI-alapú megoldások önállóan nem elegendőek, és a felhasználói viselkedés is kulcsszerepet játszik a dezinformáció elterjedésében. Az említett platformok olyan MI-alapú eszközöket alkalmaznak, amelyek célja a dezinformáció felismerése és a felhasználók figyelmeztetése: a) **automatikus figyelmeztetések,** a kétes forrásokkal kapcsolatos tartalmak alatt a platformok figyelmeztetéseket jelenítenek meg; b) **tartalom eltávolítása,** az MI-alapú algoritmusok képesek automatikusan eltávolítani azokat a tartalmakat, amelyek megsértik a közösségi irányelveket; c) **álhírek ellenőrzése**, egyes platformok harmadik fél általi tényellenőrzési programokat integrálnak, amelyek az MI-alapú szűréseket kiegészítik. A Facebook például az MI és az emberi tényellenőrzők kombinációját alkalmazza, hogy azonosítsa az álhíreket, míg a YouTube a hamis tartalmak minimalizálása érdekében algoritmusait a megbízható források előtérbe helyezésére optimalizálja (Newman et al. 2024). Bár az MI-alapú megoldások hatékonyak lehetnek, számos korlátjuk is van: a) **átláthatatlanság,** az algoritmusok döntéshozatali folyamatai gyakran nem nyilvánosak, ami bizalomvesztést okozhat a felhasználók körében (Floridi 2023); b) **túlzott szűrés,** az álhírek elleni küzdelem során előfordulhat, hogy az algoritmusok véletlenül hiteles tartalmakat is eltávolítanak, ami felveti a cenzúra problémáját (Zuboff 2019).

Egyéb lehetőségek az álhírek felismerésére és visszaszorítására

A többszintű, azaz az MI-alapú rendszerek és az emberi tényellenőrök kombinációja az egyik leghatékonyabb módszer az álhírek kezelésére. Az emberi ellenőrök feladata, hogy ellenőrizzék az MI által azonosított gyanús tartalmakat, miközben az algoritmusok a tömeges adatfeldolgozást végzik. Demeter Márton (2024) szerint ez a szinergia lehetővé teszi az álhírek gyorsabb és pontosabb felismerését.

A blokklánc-technológia, amely biztonságos és átlátható adatnyilvántartást biztosít, ígéretes megoldás lehet az információk hitelességének ellenőrzésére. Ez a technológia lehetővé teszi, hogy a hírek forrása és eredetisége nyomon követhető legyen. Fehér Katalin (2018) hangsúlyozza, hogy a blokklánc-alapú hitelesítési rendszerek jelentős szerepet játszhatnak az álhírek elleni küzdelemben, különösen az online tartalmak terén.

A mesterséges intelligencia olyan algoritmusokat is fejleszt, amelyek azonosítják az érzelmileg manipulált tartalmakat. Ezek az algoritmusok a nyelvi mintázatok és a vizuális jelek elemzésével képesek felismerni a szándékosan túlzott érzelmi töltetű vagy provokatív tartalmakat (Shu et al. 2017). Az érzelmi manipuláció korlátozása különösen fontos a társadalmi konfliktusok és a választási kampányok idején.

A technológiai megoldások mellett társadalmi összefogásra is szükség van. A média- és információs műveltségi programok, az oktatás, valamint a kormányzati és civil szervezetek együttműködése hosszú távon hozzájárulhat az álhírek terjedésének megakadályozásához. Andok Mónika (2022) úgy véli, hogy az álhírek kezelése nemcsak technológiai, hanem kulturális és társadalmi kérdés is.

Etikai dilemmák és a lehetséges megoldások, irányok

Az álhírek elleni küzdelemben a mesterséges intelligencia alkalmazása nemcsak technológiai, hanem etikai kérdéseket is felvet, amelynek megoldására integrált megközelítések szükségesek. Az átláthatóság növelése, a felelősségi kérdések tisztázása, valamint a szólásszabadság és a cenzúra közötti egyensúly megteremtése kulcsfontosságú mind az egyéni felhasználók, mind a digitális közösségek bizalmának visszaszerzése érdekében. Az etikus technológiafejlesztés, a felhasználói tudatosság növelése, a megfelelő jogi szabályozás és a nemzetközi együttműködés esszenciális a dezinformáció elleni hatékony fellépéshez.

Etikai dilemmák az álhírek elleni küzdelemben

Az MI által hozott döntések átláthatóságának hiánya komoly etikai kihívást jelent (Taddeo – Floridi 2018). A mesterséges intelligencia döntéshozatali folyamatainak átláthatósága kulcsfontosságú, különösen az álhírek elleni küzdelemben. Az MI algoritmusai gyakran működnek úgynevezett "fekete doboz"-szerű módon, ahol a felhasználók és a szabályozók nem férhetnek hozzá ahhoz, hogy pontosan megértsék, milyen paraméterek alapján születik egy döntés (Floridi 2023). Ez az átláthatóság hiánya nemcsak a technológiai vállalatok iránti bizalmat csökkenti, hanem etikai kérdéseket is felvet a felelősségvállalás terén. Például, ha egy álhírt tartalmazó bejegyzést tévesen nem szűrnek ki, ki a felelős? A technológiai cég, az MI-algoritmus készítője, vagy a felhasználó, aki az információt terjesztette? Az ilyen kérdések megválaszolása egyre sürgetőbb, ahogyan az MI-alapú rendszerek integrálódnak a mindennapi életbe (Taddeo – Floridi 2018). Az MI gyakran a tanítására használt adatok torzításait tükrözi, ami kérdéseket vet fel az objektivitásról és az igazságosságról (O'Neil 2016). Andok Mónika (2022) szerint az átláthatóság hiánya hosszú távon alááshatja a társadalmi együttműködés alapjait, mivel a technológiák használatát nem kíséri megfelelő társadalmi ellenőrzés. Floridi (2023) kiemeli azt is, hogy az átláthatóság növelése elengedhetetlen az MI-rendszerek

társadalmi elfogadottságához. Átláthatósági megoldásra irányuló javaslatok közé tartozik egyrészt az algoritmusok auditálása, azaz független szakértők bevonása az MI-rendszerek felülvizsgálatába és szabályozásába, másrészt a felhasználói visszacsatolás, ami a mesterséges intelligencia által hozott döntések magyarázatának biztosítása a felhasználók számára, pl. egyértelmű indoklások formájában. Fehér et al. (2024) tanulmánya rámutat arra, hogy a felelősség pontos meghatározása érdekében új szabályozási keretek szükségesek, amelyek figyelembe veszik a technológiai fejlődés komplexitását és társadalmi hatásait.

Az álhírek szűrésének egyik alapvető etikai kihívása a szólásszabadság és a cenzúra közötti egyensúly megteremtése. Az MI-alapú rendszerek gyakran eltávolítanak bizonyos tartalmakat a platformokról, hogy megakadályozzák a dezinformáció terjedését. Ez azonban olyan problémákhoz vezethet, mint a túlzott cenzúra: az MI tévedhet és hiteles tartalmat is eltávolíthat, vagy mint a véleményszabadság korlátozása. A demokratikus társadalmak egyik alappillére a szabad véleménynyilvánítás és Zuboff (2019) úgy véli, hogy a platformok túlzott beavatkozása alááshatja ezt az alapelvet. A túlzott cenzúra kockázatot jelent, mivel gátolhatja az egyéni véleménynyilvánítást és a nyilvános diskurzust (Zuboff 2019). Ugyanakkor a teljes szabadság megőrzése növelheti a dezinformáció terjedését, ami alááshatja a demokratikus intézményekbe vetett bizalmat.

A szólásszabadság és a platformszintű szabályozás közötti egyensúly megteremtése kiemelt jelentőségű. Egyes szakértők szerint a platformoknak nemcsak moderátori szerepet kellene betölteniük, hanem felelősséget kellene vállalniuk az általuk közvetített tartalmakért is (Wardle – Derakhshan 2017). A megoldás nem az abszolút cenzúra lehet, hanem a tartalmak átláthatóbb kezelése, ami magába foglalja a kétes tartalmak címkézését és a tényellenőrző együttműködést. Ahelyett tehát, hogy az MI eltávolítaná a tartalmakat, figyelmeztető címkék elhelyezése lehetővé tenné a felhasználók számára a kritikus értékelést és az MI és az emberi tényellenőrök együttes alkalmazása hatékonyabb lehet, mint bármelyik megoldás külön-külön.

Az MI-alapú rendszerek által végzett tartalomszűrés gyakran összeütközésbe kerül a szólásszabadság alapelvével. Az algoritmusok célja az álhírek eltávolítása, de ez nem mindig történik hibátlanul: a túlzott szűrés során a platformok néha hiteles tartalmakat is eltávolítanak, mert az algoritmusok tévesen álhírként azonosítják azokat, valamint politikai torzítás is előfordulhat, hiszen egyes rendszerek torzítva működhetnek, előnyben részesítve bizonyos nézőpontokat másokkal szemben (Wardle – Derakhshan 2017). Zuboff (2019) szerint a túlzott szűrés és a véleménynyilvánítás korlátozása alááshatja a demokratikus társadalmak alapvető értékeit. A szólásszabadság és a platformok szabályozása közötti egyensúly megteremtése tehát alapvető fontosságú.

Andok (2015) a médiahatások kapcsán hangsúlyozza, hogy a technológiai megoldások nem lehetnek teljesen függetlenek a kulturális és társadalmi kontextustól. Az MI-alapú rendszerek alkalmazása során figyelembe kell venni a helyi kulturális normákat, valamint a társadalmi bizalom szintjét. A Reuters Institute (2023) pedig rámutat, hogy a globális együttműködés elengedhetetlen az álhírek elleni küzdelemben. Az álhírek nem ismernek határokat, ezért a szabályozási és technológiai megoldásoknak nemzetközi szinten kell működniük. Fehér et al. (2024) tanulmánya szerint a nemzetközi szabályozás és az adatmegosztási együttműködés növelheti az MI hatékonyságát az álhírek elleni küzdelemben.

Az MI és az álhírek jövője: lehetséges megoldások és irányok

Az álhírek elleni küzdelemben az MI szerepe várhatóan tovább növekszik és felmerül az etikus technológiafejlesztés szükségessége is. A mesterséges intelligencia fejlesztése során elengedhetetlen ugyanis az etikai szempontok figyelembevétele. Ez magában foglalja az algoritmusok átláthatóságának növelését, valamint olyan szabványok kidolgozását, amelyek

biztosítják az MI alkalmazásának társadalmilag felelős módját (O'Neil 2016). Az MI-rendszereknek átláthatóvá kell tenniük a döntéshozatali folyamatokat (Floridi 2023). Továbbá felelősségi protokollok kialakítása is mérvadó, valamint a platformok és fejlesztők számára egyértelmű felelősségi irányelvek megalkotása.

A felhasználók kritikai gondolkodásának fejlesztése is lényeges elem a folyamatban, nem csupán a technológiai aspektus. Pennycook és Rand (2019: 25–28) azt vallják, hogy a médiaműveltség javítása, a források kritikus értékelése és az álhírek felismerésére irányuló oktatási programok jelentős mértékben csökkenthetik az álhírek terjedésének hatékonyságát. Ezeket a programokat kiegészítheti az MI-alapú eszközök használatának ismertetése annak érdekében, hogy a felhasználók jobban megértsék a tartalomszűrő algoritmusok működését. A médiaműveltség szerepe ezúton is felértékelődik: a felhasználók oktatása, az információs műveltség fejlesztése, amely magában foglalja az álhírek felismerésének képességét és az információforrások kritikus értékelését, hozzájárulhat a dezinformáció hatásainak csökkentéséhez (Aczél 2015). A digitális kompetenciák, az információforrások hitelességének értékelése és az információs műveltség beépítése, valamint a tudatos médiahasználatra irányuló kampányok szervezése akár iskolai programokon, akár felnőttképzésen keresztül is javasolt.

Az álhírek globális problémát jelentenek, amelynek megoldása nem korlátozódik egyetlen ország vagy régió határain belülre, ezért a nemzetközi szabályozás és együttműködés létfontosságú továbbá a kérdéskör hatékony kezeléséhez. A 2024-es Digital News Report például hangsúlyozza, hogy a technológiai vállalatoknak, civil szervezeteknek és a szabályozóknak szorosabban együtt kell működniük annak érdekében, hogy átláthatóbb és biztonságosabb digitális környezetet teremtsenek, pl. egységes szabványokat hozzanak létre a platformok felelősségének és átláthatóságának növelése érdekében, valamint közös adatbázisokat fejlesszenek az álhírek azonosításához szükséges információk megosztása érdekében. A globális szabályozás és a technológiai innovációk összehangolása alapvető fontosságú. Az elemzés szerint az MI-alapú rendszerek hatékonysága növelhető, ha a technológiai vállalatok és a szabályozók szorosabb együttműködést alakítanak ki. Az MI alkalmazása tehát az álhírek elleni küzdelemben hatékony eszköz, de csak akkor lehet valóban sikeres, ha technológiai, társadalmi és szabályozási szempontokat egyaránt figyelembe veszünk. Az átláthatóság növelése, a felhasználók tudatosságának fejlesztése, valamint a nemzetközi együttműködés mind-mind alapvető lépés a hiteles digitális környezet megteremtése érdekében.

Összefoglalás

Az álhírek terjedésében és visszaszorításában a mesterséges intelligencia kettős szerepet játszik, miközben számos etikai és társadalmi kérdést vet fel. Miközben az MI-eszközök jelentős előrelépést jelentenek a dezinformáció azonosítása és szűrése terén, ezek a technológiák önmagukban nem nyújtanak teljes körű megoldást. A társadalmi szemléletformálás, a felhasználók oktatása, a szabályozási keretek megalkotása és a technológiai vállalatok felelősségteljes működése nélkülözhetetlenek a hiteles információáramlás fenntartásához.

Az MI által támogatott technológiák, mint például a chatbotok, a deepfake videók és a szöveggeneráló algoritmusok, radikálisan növelték a dezinformáció előállításának és terjesztésének hatékonyságát. A közösségi média algoritmusai tovább fokozzák ezt a hatást azáltal, hogy a felhasználók viselkedési mintái alapján személyre szabott tartalmakat kínálnak, amelyek gyakran előnyben részesítik az érzelmileg túlfűtött, polarizáló üzeneteket. Az ilyen technológiák nemcsak az igazság és hitelesség fogalmát kérdőjelezik meg, hanem komoly fenyegetést jelentenek a demokratikus nyilvánosságra is.

A mesterséges intelligencia ugyanakkor kulcsszerepet játszhat az álhírek felismerésében és visszaszorításában. A természetes nyelvfeldolgozás (NLP) és a gépi tanulási algoritmusok képesek azonosítani a dezinformációs mintázatokat, például a hiteltelen forrásokat vagy a manipulatív narratívákat. A közösségi média platformok és keresőmotorok MI-alapú tartalomszűrő rendszerei jelentős előrelépést jelentenek az álhírek terjedésének megakadályozásában. Ugyanakkor ezek az eszközök gyakran átláthatatlan módon működnek, ami bizalomvesztést okozhat a felhasználók körében.

Az MI alkalmazása az álhírek elleni küzdelemben számos etikai kérdést vet fel. Az átláthatóság hiánya és a „fekete doboz”-szerű működés különösen aggasztó, mivel a felhasználók és a szabályozók nem érthetik meg teljes mértékben az algoritmusok működését. Ez alááshatja a technológiai vállalatok iránti bizalmat, és felveti a felelősség kérdését is: ki tehető felelőssé, ha egy álhírt nem szűrnek ki, vagy ha egy hiteles tartalmat tévesen eltávolítanak? A szólásszabadság és a cenzúra közötti egyensúly megteremtése szintén alapvető etikai kihívás, hiszen a túlzott szűrés alááshatja a véleménynyilvánítás szabadságát, míg a túlzott szabadság növelheti a dezinformáció terjedésének veszélyét.

A tanulmány a jövőbeli megoldási lehetőségekre is rávilágított. Az etikus technológiafejlesztés, az algoritmusok átláthatóságának növelése, valamint a szabályozási keretek kidolgozása alapvető lépések a dezinformáció elleni küzdelemben. A felhasználói médiaműveltség fejlesztése szintén kulcsfontosságú, hiszen az álhírek elleni védekezés nemcsak technológiai, hanem társadalmi felelősség is. A nemzetközi együttműködés elengedhetetlen, mivel az álhírek globális problémát jelentenek, amely csak összehangolt erőfeszítésekkel kezelhető hatékonyan.

Összességében elmondható, hogy az MI-alapú technológiák egyszerre jelentenek kihívást és megoldást az álhírek problémájára. A technológia önmagában nem elegendő; szükség van a társadalmi, kulturális és szabályozási megközelítések integrálására, hogy egy átláthatóbb, biztonságosabb és hitelesebb digitális környezetet teremtsünk. Az álhírek elleni küzdelem nemcsak az MI jövőjét, hanem a demokratikus társadalmak alapvető értékeit is meghatározza.

A jövő kulcsa tehát az átláthatóságban, az együttműködésben és az információs műveltség fejlesztésében rejlik. Az álhírek elleni küzdelem nemcsak technológiai, hanem társadalmi szinten is kihívást jelent, amely az összes érintett fél együttműködését igényli.

Irodalom

Aczél Petra (2017) Az álhír. Kommentár a jelenség értelmezéséhez. *Századvég*, 84. 5–25.

Aczél Petra – Veszelszki Ágnes (szerk.) (2023) *Deepfake: A valótlan valóság*. Budapest, Gondolat Kiadó.

Aczél Petra (2015) Médiaműveltség. In: Aczél Petra (2015 szerk.) *Műveljük a médiát!* Budapest, Wolters Kluwer. 133–177.

Allcott, Hunt – Gentzkow, Matthew (2017) Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31/2. 211–236. https://doi.org/10.1257/jep.31.2.211

Andok Mónika (2015) Médiahatások. In: Aczél Petra (2015 szerk.) *Műveljük a médiát!* Budapest, Wolters Kluwer. 97–132.

Andok Mónika (2022) *Digitális vallás: Az új információs technológiák hatása a vallásokra, vallási közösségekre Magyarországon*. Budapest, Ludovika Egyetemi Kiadó.

Brynjolfsson, Erik – McAfee, Andrew (2014**)** *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, W.W. Norton & Company.

Falyuna Nóra – Fehér Katalin – Demeter Márton – Szüdi Gábor – Uscinski, Joseph E. **(2024)** *Against Disinformation: Bridging Science and Public Discourse*. *Információs Társadalom*, 24/2. 68–84. https://doi.org/10.22503/inftars.XXIV.2024.2.4

Demeter Márton (2019) The World-Systemic Dynamics of Knowledge Production: The Distribution of Transnational Academic Capital in the Social Sciences. *Journal of World-Systems Research*, 25/1. 111–144. https://doi.org/10.5195/jwsr.2019.887

DiResta, Renee – Shaffer, Kris – Ruppel, Becky – Sullivan, David – Matney, Robert – Fox, Ryan – Albright, Jonathan – Johnson, Ben (2018) *The Tactics & Tropes of the Internet Research Agency*. Austin, New Knowledge.

Fehér Katalin (2018) Digital Realities: The Impact of Digital Technologies on Reality Perception. *Információs Társadalom*, 18/4. 5–18.

Fehér Katalin – Király Olívia (2017) Álhíresülés - A hamis hírek dinamikája a médiában. Századvég, 84. 39–48.

Fehér Katalin – Vicsek Lilla – Deuze, Mark (2024) *Modeling AI Trust for 2050: Perspectives from media and info-communication experts*. *AI & Society*, 2933–2946. https://doi.org/10.1007/s00146-023-01827-6

Floridi, Luciano **(2023)** *The Ethics of Artificial Intelligence*: Principles, Challenges, and Opportunities. Oxford, Oxford University Press. https://doi.org/10.1093/oso/9780198883098.001.0001

Newman, Nic – Fletcher, Richard – Eddy, Kirsten – Robertson, Craig T. – Nielsen, Rasmus Kleis (2023) *Reuters Institute Digital News Report 2023*. Oxford, Reuters Institute for the Study of Journalism, University of Oxford. https://doi.org/10.60625/risj-p6es-hb13

Newman, Nic – Fletcher, Richard – Robertson, Craig T. – Arguedas, Amy Ross – Nielsen, Rasmus Kleis (2024) *Reuters Institute Digital News Report 2024*. Oxford, Reuters Institute for the Study of Journalism, University of Oxford. https://doi.org/10.60625/risj-vy6n-4v57

O'Neil, Cathy (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, Crown Publishing Group.

Pariser, Eli (2011) *The Filter Bubble*. New York, Penguin Press.

Pennycook, Gordon – Rand, David G. (2019) Fighting misinformation on social media using crowdsourced judgments of news source quality. PNAS, 116/7. 2521–2526. https://doi.org/10.1073/pnas.1806781116

Shu, Kai – Sliva, Amy – Wang, Suhang – Tang, Jiliang – Liu, Huan (2017***)*** *Fake News Detection on Social Media: A Data Mining Perspective*. *ACM SIGKDD Explorations Newsletter*, 19/1. 22–36. https://doi.org/10.1145/3137597.3137600

Taddeo, Mariarosaria – Floridi, Luciano (2018) How AI can be a force for good. *Science*, 361/6404. 751–752. https://doi.org/10.1126/science.aat5991

Vosoughi, Soroush – Roy, Deb – Aral, Sinan (2018) The spread of true and false news online. *Science*, 359/6380. 1146–1151. https://doi.org/10.1126/science.aap9559

Wardle, Claire – Derakhshan, Hossein (2017) Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Strasbourg, Council of Europe.

Zuboff, Shoshana (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, PublicAffairs.