Iustum Aequum Salutare
XXI. 2025. 3. • 61–78.

A MESTERSÉGES INTELLIGENCIA SZÓLÁSSZABADSÁGA ÉS TARTALOMMODERÁLÁSI GYAKORLATA A KÖZÖSSÉGI MÉDIÁBAN A DSA ÉS AZ MI RENDELET VONATKOZÁSÁBAN

Gosztonyi Gergely[*] – Gyetván Dorina[**] – Kovács Andrea[***]

1. Bevezetés

Bár Gartner technológiák felfutását vizsgáló úgynevezett Hype Cycle görbéjén a generatív mesterséges intelligencia már 2024 augusztusában a 'csalódási fázisba' került,[1] egyre több szoftverben tapasztaljuk valamilyen mesterséges intelligencia megjelenését. Ez a trend nem kerülheti el a közösségi média platformokat sem, a Facebook és az Instagram mögött álló Meta például saját mesterséges intelligencia modellt fejleszt Llama néven, amelyet a felhasználói bátran testre szabhatnak.[2] Emellett a Meta nyíltan vállaltan használ mesterséges intelligenciát a tartalmak szűrésére,[3] ahogyan az X / Twitter, valamint a TikTok is hasonló módszert alkalmaz.

A mesterséges intelligencia – a közvélemény számára hirtelennek tűnő – felkapottsága a jogalkotókat sem hagyta hidegen. A probléma ugyanakkor az, hogy – Pázmándi Kinga szavaival – „a tolómérce egyik végén a jog által biztosítani kívánt védelmi célok […] állnak, a másik végén a társadalmilag jól hasznosuló innováció, a

[*] Egyetemi tanár, Eötvös Loránd Tudományegyetem, Állam- és Jogtudományi Kar, gosztonyi@ajk.elte.hu, ORCID: https://orcid.org/0000-0002-6551-1536

[**] PhD hallgató, Eötvös Loránd Tudományegyetem Állam- és Jogtudományi Doktori Iskola, dorina.gyetvan@gmail.com, ORCID: https://orcid.org/0000-0001-5361-0011

[***] PhD hallgató, Eötvös Loránd Tudományegyetem Állam- és Jogtudományi Doktori Iskola, ankovacs@student.elte.hu, ORCID: https://orcid.org/0000-0003-4970-3551

1 2024 Hype Cycle for Emerging Technologies Highlights Developer Productivity, Total Experience, AI and Security. *Gartner,* 2024. szeptember 21. https://tinyurl.com/5n8jx2uz

2 https://www.llama.com

3 Meta: Meta's New AI System to Help Tackle Harmful Content. *Meta,* 2021. december 8. https://tinyurl.com/4w9xup4r

https://doi.org/10.69695/ias.2025.3.05

»tolómérce értékeinek beállítását« pedig állandó […] bizonytalanság kíséri."[4] A tanulmány megírásának időpontjában az elmúlt egy éves időtartamban az Európai Unióban (a továbbiakban: EU) elfogadták a mesterséges intelligencia rendeletet[5] (a továbbiakban: MI rendelet) és hozzá kapcsolódóan a felelősségi irányelv tervezetet,[6] az Egyesült Államok korábbi elnöke, Joseph Biden pedig kibocsátott egy utasítást a biztonságos és megbízható mesterséges intelligencia fejlesztéséről és használatáról (a továbbiakban: E.O. 14110).[7] Az utasítást Donald J. Trump a mesterséges intelligencia fejlesztés akadályának tekintette és több más korábbi elnöki rendelettel együtt 2025 elején visszavonta,[8] egyúttal utasítva az érintett szervezeteket, hogy azonnali hatállyal vizsgálják felül korábbi lépéseiket.[9]

Bár általánosan elfogadott definíció nem létezik a szakirodalomban,[10] mind az EU, mind pedig az Egyesült Államok meghatározza, hogy mit ért mesterséges intelligencia alatt. Az EU rendelete szerint a mesterséges intelligencia olyan szoftver, amely mögött meghatározott technológia áll (gépi tanulás) vagy meghatározott megközelítéssel került kifejlesztésre (logikai és tudás alapú, illetve statisztikai), és amely képes a környezetét befolyásolni.[11] Az E.O. 14110 és az azt váltó elnöki rendelet, a US Code definíciójára hivatkozik, mely szerint a mesterséges intelligencia egy gépi rendszer, amely előrejelzéseket ad, ajánlásokat tesz vagy döntéseket hoz, amelyhez a bemeneti adatokat

---

4 Pázmándi Kinga: Digitalizáció, technológiai fejlődés, jogi paradigmák. *Gazdaság és Jog,* 2018/12. 11. https://doi.org/10.21637/GT.2018.01.02

5 Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról (a mesterséges intelligenciáról szóló rendelet), PE/24/2024/REV/1, HL L, 2024/1689, 2024.7.12. Varju Márton: A mesterséges intelligencia szabályozása az Európai Unióban: szakpolitika, szabályozási stratégia és jog. In: Mezei Kitti (szerk.) *Kockázatok és lehetőségek a mesterséges intelligencia jogi szabályozásában*. Budapest, Wolters Kluwer, 2025. 33–50.

6 Javaslat az Európai Parlament és a Tanács irányelve a szerződésen kívüli polgári jogi felelősségre vonatkozó szabályoknak a mesterséges intelligenciához való hozzáigazításáról (a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv) COM(2022) 496 final 2022/0303 (COD). Ki kell ugyanakkor emelni, hogy 2025 tavaszán az Európai Bizottság *ad acta* kívánja helyezni a felelősségi irányelv tervezetét. Isabel Marques da Silva: Vajon az MI-specifikus felelősségi szabályok visszavonása jogorvoslat nélkül hagyja a károsultakat? *Euronews,* 2025. március 4. https://tinyurl.com/2xnkt38f

7 Executive Order (E.O.) 14110 on Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, 88 FR 75191.

8 The White House: Initial Rescissions of Harmful Executive Orders and Actions. *whitehouse.gov,* 2025. január 20. https://tinyurl.com/4wuhbe8h

9 The White House: Removing Barriers to American Leadership in Artificial Intelligence. *whitehouse.gov,* 2025. január 23. https://tinyurl.com/2u6zf73x Sec. 5.

10 Necz Dániel: A mesterséges intelligencia felhasználásával történő adatkezelések egyes sajátos szempontjai. *Acta Humana,* 2023/3. 98. https://doi.org/10.32566/ah.2022.3.4; Ambrus István: A mesterséges intelligencia és a büntetőjog. *Állam- és Jogtudomány,* 2020/4. 6–9.

11 MI rendelet 3. cikk 1.

modellekké absztrahálja.[12] Megjegyzendő, hogy a definíció eredeti kontextusát tekintve a kereskedelmi szabályok között helyezkedik el a National AI Initiative cím alatt.

A jogalkotókkal szemben a szakirodalom már szentimentálisabb, sokkal inkább az emberi gondolkodáshoz való hasonlóságra vagy az attól való különbségekre koncentrál.[13] Amikor a szakirodalom a mesterséges intelligencia szólásszabadságát – és ehhez kapcsolódóan a személyiség kérdését – vizsgálja, akkor általában a kevésbé technikai, sokkal inkább emberi vonások kerülnek előtérbe, amellyel jelen tanulmány első része foglalkozik. A tanulmány a továbbiakban vizsgálja a tartalommoderálás általános kérdéseit és szabályozását az EU-ban, illetve a mesterséges intelligencia – pozitív és negatív – szerepét és esélyeit a tartalommoderálási gyakorlatban.

2. A mesterséges intelligencia szólásszabadsága

Egyes mesterséges intelligenciák esetében láthattunk már példát arra, hogy alkotójától elkülönülten valamilyen joggal ruházzák fel, mint Sophia[14] és Mirai esetében.[15] A mesterséges intelligencia egyre elterjedtebb alkalmazása miatt azonban leggyakrabban a felelősségi kérdésekkel foglalkozik a szakirodalom,[16] ugyanakkor – amennyiben a fizikai dimenziótól eltekintünk – a robotjog már régóta foglalkozik az ember által megteremtett, mesterségesen létrehozott, intelligens gépek lehetséges jogaival, amelyekbe beletartozik a kommunikációhoz való jog.[17]

A mesterséges intelligencia és a szólásszabadság kapcsán érdemes megvizsgálni, hogy ki a szólásszabadság alanya. Potenciális alany lehet:

(i) a mesterséges intelligencia, vagy
(ii) egy mögöttes alany, közülük pedig

---

12 15 USC 9401: Definitions (Pub. L. 116–283, div. E, §5002, Jan. 1, 2021, 134 Stat. 4523.); E.O. 14110 Section 3. (b). Theodore S. Boone: An examination of certain key features of the new White House Executive Order on Artificial Intelligence. *Corvinus Law Papers,* 2024/1. 1–19.

13 Az erős és gyenge mesterséges intelligenciák jellemzőiről ld. összefoglalóan: Kovács Andrea: A mesterséges intelligencia jogi megítélésének egyes kérdései. *Themis,* 2024/1. 37–40. https://doi.org/10.55052/themis.2024.1.36

14 Sophia például állampolgárságot kapott. Molnár Balázs: Személy-e a mesterséges intelligencia? In: Varga János – Csiszárik-Kocsir Ágnes – Garai-Fodor Mónika (szerk.): *Vállalkozásfejlesztés a XXI. században 2023. II. kötet. A jelen kor gazdasági kihívásainak és társadalmi változásainak interdiszciplináris megközelítései*. Budapest, Óbudai Egyetem Keleti Károly Gazdasági Kar, 2023. 280–283.

15 Mirai az üzemeltető vállalattól függetlenül letelepedési engedélyt kapott. Nagy Teodóra: A jövő kihívásai: robotok és mesterséges intelligencia az alapjogi jogalanyiság tükrében. *MTA Law Working Papers,* 2020/6. 13.

16 Ld. pl. Stefán Ibolya: A mesterségesintelligencia-rendszerek felelősségi kérdései az uniós dokumentumok és a magyar szabályozás tükrében. *Publicationes Universitatis Miskolcinensis, Sectio Juridica et Politica,* 2022/2. 364–387. https://doi.org/10.32978/sjp.2022.030; Csitei Béla: Az önvezető gépjárművek és a polgári jogi kárfelelősség. In: Glavanits Judit (szerk.): *A gazdasági jogalkotás aktuális kérdései*. Budapest, Dialóg Campus, 2019. 25–26.

17 Yurii Vadymovych Sheliazhenko: Artificial Personal Autonomy and Concept of Robot Rights. *cyberleninka.ru,* 2017. április 12., 17–21. http://dx.doi.org/10.20534/EJLPS-17-1-17-21 ; Klein Tamás: Robotjog. In: Klein Tamás – Tóth András (szerk.): *Technológiajog – Robotjog – Cyberjog*. Budapest, Wolters Kluwer Kft., 2018. 179–215. https://doi.org/10.55413/9789632958293

a) a programozó,
b) az üzemben tartó / üzemeltető, vagy
c) a használó.

A lehetséges alanyok közül maga a mesterséges intelligencia alanyisága a legvitatottabb. A szólásszabadság alanyaként való elismerésének lehetősége nem idegen a szakirodalomtól,[18] ehhez különböző szempontokat határoznak meg az egyes szerzők, amelyek az egészen technikai, viszonylag egyszerűen meghatározható feltételektől az absztrakt feltételekig terjednek. Egyszerűen meghatározható szempont például, hogy a szólásszabadság elismeréséhez elegendő az is, hogy a kimeneti eredmény indeterminált, nem megjósolható.[19] Hasonló álláspontot képvisel Tim Wu is, aki szerint akkor beszélhetünk személyiségről, ha (1) képes koncepcionális gondolkodásra és (2) képes kifejezni a kialakult véleményét.[20] Haladva az absztraktabb feltételek felé, Sheliazhenko szerint a szólásszabadság elismeréséhez azt szükséges vizsgálni, hogy rendelkezik-e a vizsgálat tárgya olyan társadalmilag komplex funkcionalitással, amely az emberi autonómia határait súrolja,[21] Schwithgebel és Garza szerint pedig a megfelelő szociális státusz is elegendő lehet.[22] Érdemes azonban megemlíteni, hogy a mesterséges intelligencia is felruházható jogi személyiséggel – így bizonyos jogokkal – is, a jogi személyiség megadása pedig a jogalkotó diszkrecionális döntése és mint ilyen, akár praktikussági szempontokat is követhet.[23]

Más szemszögből nézve azonban a mesterséges intelligencia nem önmagában való, mindig áll mögötte valamilyen emberi tevékenység,[24] így amennyiben nem a mesterséges intelligenciának tulajdonítjuk a szólásszabadságot, úgy a mögöttes személyeket szükséges vizsgálnunk.[25] A programozó, az üzemben tartó/üzemeltető és a használó

---

18 Bert-Jaap Koops – Mireille Hildebrandt – David-Olivier Jaquet-Chiffelle: Bridging the Accountability Gap: Rights for New Entities in the Information Society? *SSRN,* 2010. július 23. 555–559. *Minnesota Journal of Law, Science & Technology,* Vol. 11., No. 2. (2010) 497–561.

19 Marek Świerczyński – Zbigniew Więckowski: Statut Jednolity Sztucznej Inteligencji. *Zeszyty Prawnicze,* Vol. 23., No. 1. (2023) 220. https://doi.org/10.21697/zp.2023.23.1.09

20 Tim Wu: Machine Speech. *University of Pennsylvania Law Review,* Vol. 161., No. 6. (2013) 1495., 1503. Ezzel kapcsolatban ld. még: *Autronic AG v Switzerland*, no. 12726/87 1990. május 22-i ítélet, 47.

21 Sheliazhenko i. m. 19.

22 Eric Schwitzgebel – Mara Garza: A Defense of the Rights of Artificial Intelligences. *Midwest Studies In Philosophy,* Vol. 39., No. 1. (2015) 102–103. https://doi.org/10.1111/misp.12032

23 Beatriz A. Ribeiro et al.: Metacognition, Accountability and Legal Personhood of AI. In: Henrique Sousa Antunes et al. (szerk.): *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Cham, Springer, 2024. 174., 181–182. https://doi.org/10.1007/978-3-031-41264-6_9; Visa Aj Kurki: *A Theory of Legal Personhood*. Oxford, Oxford University Press, 2019. 177., 179. https://doi.org/10.1093/oso/9780198844037.001.0001; Keserű Barna Arnold: A mesterséges intelligencia néhány magánjogi aspektusáról. In: Glavanits Judit (szerk.): *A gazdasági jogalkotás aktuális kérdései*. Budapest, Dialóg Campus, 2019. 116. A témáról bővebben ld. Kovács i. m.

24 Margot E. Kaminski – Meg Leta Jones: Constructing AI Speech. *The Yale Law Journal Forum,* Vol. 133., No. 1. (2024) 1226.

25 Federico Gustavo Pizzetti: Embryos, Organoids and Robots: 'legal subjects'? *BioLaw Journal,* Vol. 9., No. 1. (2021) 348–349. https://doi.org/10.15168/2284-4503-755; Matthew Hines: I Smell a Bot:

esetében ezek a szerepkörök átjárhatók, közöttük nem húzódik éles határ. Moderálás tekintetében példákkal szemléltetve a lehetséges eseteket:

a) a programozó, az üzemben tartó/üzemeltető és a használó megegyezik. Ebben az esetben ugyanaz a vállalat programozza, üzemelteti (elhárítja a hibákat, 1-2-n szintű támogatást nyújt) és használja (promptolja, paraméterezi, döntéseket hoz a kimenet alapján);
b) a vállalat a mesterséges intelligencia kapcsán egy másik vállalatot bíz meg a fejlesztéssel, de azt már ő maga üzemelteti és használja;
c) a cég csak megvásárolja és használja a mesterséges intelligenciát, míg a programozást és az üzemeltetést más végzi.

Míg az a) esetben a helyzet egyértelmű, a jogok és a felelősség is a vizsgált vállalathoz kapcsolódik, addig a c) esetben már az is kérdés, hogy mennyiben egy kompakt, úgynevezett dobozos szoftverről, és mennyiben egyedi fejlesztésű megoldásról van szó. Az üzemeltető szerepkörét tekintve iránymutatásként használható az MI rendelet, amely e szerepkörebe tartozónak tekinti a szolgáltatót, az alkalmazót, a meghatalmazott képviselőt, az importőrt és a forgalmazót is, amennyiben a mesterséges intelligencia veszélyes üzem.[26]

Ezen szereplők közül azonban feltehetően azt fogjuk beszélőnek tekinteni, aki a tartalom generálásához (vagy jelen esetben moderálásához) szükséges paramétereket meghatározta.[27] Amiben azonban a programozó, az üzemeltető és használó szerepkörre megegyezik, hogy természetes és/vagy jogi személyiséggel rendelkeznek, amelyek rendelkeznek szólásszabadsággal. A közösségi média szolgáltatók tekintetében a saját fejlesztés miatt a mesterséges intelligencia tevékenysége feltehetően a szolgáltatóknak lesz betudható, különösen arra tekintettel, hogy azokat a szabályzatokat és irányelveket is a szolgáltató írja, amelyeket végre kell hajtani.[28] Ezt az álláspontot látszik követni a 3rd Circuit ítélete az Anderson v. TikTok ügyben.[29] Szemben az alsóbb fokú bíróság ítéletével bíró értelmezésében – a Moody v. NetChoice[30] Legfelső Bírósági ítéletre alapozva – a TikTok felelősségre vonható az algoritmusa által való ajánlásért és összeállításért, mint saját kifejező produktumáért.[31] Kiemelendő azonban, hogy az ügyben végig algoritmusról van szó, nem pedig mesterséges intelligenciáról. A Moody v. NetChoice-ítélet ugyan felveti a mesterséges intelligencia alkalmazásának kérdését,

California's S.B. 1001, Free Speech, and the Future of Bot Regulation. *Houston Law Review,* Vol. 57., No. 2. (2019) 408.

26 MI rendelet 3. cikk, 3–7.

27 Hines i. m. 420.

28 Kylie Robison: Inside the Shifting Plan at Elon Musk's X to Build a New Team and Police a Platform 'so Toxic It's Almost Unrecognizable'. *Fortune,* 2024. február 7. https://tinyurl.com/37x2hu3c

29 Anderson v. TikTok Inc, No. 22-3061 (3d Cir. 2024).

30 Moody v. NetChoice, LLC. 603 U.S. 707 (2024).

31 Anderson v. TikTok, 184–186.

de mindössze annyi megjegyzést tesz, hogy ilyen eszközök alkalmazása alkotmányos jelentőségű lehet.[32]

Más szerzők azonban nem foglalkoznak a jogalanyiság kérdésével, más igazolást választanak a mesterséges intelligencia kimenetének védelméhez, részben azért, mert a mesterséges intelligencia esetében – épp annak kiszámíthatatlansága miatt – előfordulhat, hogy tevékenysége nem embernek betudható, nem vezethető vissza könnyedén emberi tevékenységre.[33] Ezen érvek többek között:

1. Nincs olyan doktrína, amely kifejezetten kizárná a mesterséges intelligenciát az Első Alkotmánykiegészítés hatálya alól.[34]
2. A mesterséges intelligencia kimenete tartalmazhat olyan elemeket, amelyek az emberek számára érdekesek, fontosak, vagy tartalmazhat olyan közléseket, amelyeket egyébként a szólásszabadság véd.[35] Ezáltal pedig ezek részt vesznek a demokratikus vitában és az információk terjesztésében.[36] Ezen érv részeként kiemelendő a hallgatóság joga az információhoz jutáshoz, amely a szólásszabadság része.[37]

A mesterséges intelligencia által generált üzenetek esetében következményekkel jár az is, ha elfogadjuk, hogy a szólásszabadság hatálya alá tartozik, és az is, ha nem. Ha nem tartozik a hatálya alá pusztán azért, mert nem közvetlenül embertől származik, akkor azzal egyébként védett beszéd cenzúrázható.[38] Ha viszont nem zárható ki a szólásszabadság hatálya alól, akkor problémát okozhat az egyébként a beszélő szándékától függően nem védett beszéd megítélése és kérdésessé válik a felelősségre vonás.[39] A közösségi média esetében – amennyiben ténylegesen saját fejlesztésű mesterséges intelligenciát használnak tartalommoderálásra – a helyzet egyértelmű: a fejlesztés és a paraméterezés (végrehajtandó szabályzatok megalkotása és a végrehajtás módja) is egy kézben van.[40]

---

32 Moody v. NetChoice.

33 Lynne Higby: Navigating the Speech Rights of Autonomous Robots in a Sea of Legal Uncertainty. *Journal of Technology Law & Policy,* Vol. 26., No. 1. (2021) 33.

34 Toni Massaro – Helen Norton: Siri-Ously? Free Speech Rights and Artificial Intelligence. *Northwestern University Law Review,* Vol. 110., No. 5. (2016) 1185.

35 Koops–Hildebrandt–Jaquet-Chiffelle i. m. 555–559.; Massaro–Norton i. m. 1192.; James Garvey: Let's Get Real: Weak Artificial Intelligence Has Free Speech Rights. *Fordham Law Review,* Vol. 91., No. 3. (2022) 953., 991.

36 Kaminski–Jones i. m. 1231–1242. A szerzők azt is kifejtik, hogy vannak olyan esetek, amikor fontos szerepet tölt be a beszélő személye, illetve szándéka (például rágalmazás esetén). A magyar Alkotmánybíróság gyakorlatában a nyilvános társadalmi kommunikációban való részvételben látja a szólásszabadság hatályát. Török Bernát: *Szabadon szólni, demokráciában. A szólásszabadság magyar doktrínája az amerikai jogirodalom tükrében.* Budapest, HVG-ORAC, 2018.

37 Emberi Jogok Európai Egyezménye 10. cikk.

38 Higby i. m. 47.; Wu i. m. 1521.

39 Higby i. m. 48.

40 Robison i. m.

3. A tartalommoderálás általános kérdései és szabályozása az EU-ban

3.1. A tartalommoderálás és a DSA, különös tekintettel az átláthatósági követelményekre

2021 márciusában Mark Zuckerberg úgy nyilatkozott kongresszusi meghallgatása során, hogy a gyűlöletbeszédet tartalmazó tartalmak 95%-át, míg – legjobb tudomása szerint – a terrorista tartalmat megtestesítő tartalmak 98-99%-át már mesterséges intelligencia azonosítja, és nem emberek.[41] A Facebook, az X/Twitter és a többi közösségi média platform moderálási gyakorlatának, illetve elveinek szabályozása világszerte nagy kihívást jelent. A digitális szolgáltatások egységes piacáról szóló 2022/2065 rendelet[42] (a továbbiakban: DSA) elsősorban a közvetítő szolgáltatókra, így az online platformokra vonatkozó szabályozást tartalmaz, azonban a tartalommoderálásra vonatkozó rendelkezések, illetve az algoritmikus átláthatósági és elszámoltathatósági követelmények folytán kiegészíti a többi uniós, mesterséges intelligencia szabályozására irányuló törekvést.[43]

A DSA e körben választ kísérel adni a tartalommoderálás egyes problémás kérdéseire is, célkitűzései között szerepel az átláthatóság és elszámoltathatóság biztosítása,[44] illetve az online platformot üzemeltető szolgáltatók tartalommoderálási döntéseinek hatékony felülvizsgálhatósága.[45]

Ugyan a DSA már meghatározza a tartalommoderálás fogalmát, a tartalommoderálásnak korábban nem volt – és a szakirodalmat tekintve továbbra sincs – tudományos konszenzuson alapuló meghatározása, az inkább egyfajta gyűjtőfogalomként fogható fel.[46] Összefoglalóan – tág értelemben – úgy határozhatjuk meg, mint amely magában foglalja a közösségi média szolgáltatók minden alapjogi kérdéseket felvető olyan eljárását, amely minden esetben felhasználói tartalmat érint az ahhoz való hozzáférhetőség emberi erőforrás vagy mesterséges intelligencia általi korlátozásával, amelynek azonban nem előfeltétele az érintett tartalom jogellenessége – mivel elsődleges „jogalapját" a jogszabályok helyett a platformok szerződéses feltételei között kell keresni –, és ame-

---

41 House Energy and Commerce Subcommittee on Communications and Technology Disinformation Nation: *Social Media's Role in Promoting Extremism and Misinformation*. 117th Congress. Testimony of Mark Zuckerberg, 2021. március 25. https://tinyurl.com/52by7xzn

42 Az Európai Parlament és a Tanács (EU) 2022/2065 rendelete (2022. október 19.) a digitális szolgáltatások egységes piacáról és a 2000/31/EK irányelv módosításáról (digitális szolgáltatásokról szóló rendelet), PE/30/2022/REV/1, HL L 277., 2022.10.27., 1–102.

43 Mayer Brown: EU Digital Services Act's Effects on Algorithmic Transparency and Accountability. *Lexology*, 2023. március 27., https://tinyurl.com/8wsa82x2

44 DSA 49. preambulumbekezdés.

45 DSA 44. és 109. preambulumbekezdés.

46 Gosztonyi Gergely: *Censorship from Plato to Social Media. The Complexity of Social Media's Content Regulation and Moderation Practices*. Cham, Springer, 2023. 7–19. https://doi.org/10.1007/978-3-031-46529-1_2

lyet a platform általában a saját gazdasági vagy társadalmi agendája[47] vagy a közvetítő szolgáltatói felelősség alóli mentesülés érdekében alkalmaz, így hiányzik a mérlegelés és a döntéshozatal átláthatósága. A motivációk közül a saját gazdasági érdekből történő moderálás (azért, hogy a felhasználóik ne találkozzanak sértő, felháborító vagy zavaró tartalmakkal) annál meghatározóbb, minél gyakoribbak a felhasználók közötti véleménycserék, és minél élénkebb a kommunikáció.[48]

Ezzel szemben a DSA szerinti tartalommoderálás a közvetítő szolgáltató olyan automatizált vagy nem automatizált tevékenysége, amely különösen a szolgáltatás igénybe vevője által közzétett jogellenes tartalom vagy a közvetítő szolgáltató szerződési feltételeivel összeegyeztethetetlen információ észlelésére, azonosítására és kezelésére szolgál, ideértve az ilyen jogellenes tartalom vagy információ elérhetőségét, láthatóságát és hozzáférhetőségét érintő intézkedéseket, például annak hátra sorolását, demonetizálását, az ahhoz való hozzáférés megszüntetését vagy annak eltávolítását, vagy a szolgáltatás igénybe vevője általi információközlés lehetőségét érintő intézkedéseket, például a fiókja megszüntetését vagy felfüggesztését.[49] A fenti két meghatározást összevetve láthatjuk, hogy nincs teljes tartalmi azonosság a két meghatározás között. A DSA nem teszi fogalmi elemmé az alapjogi vonatkozást, holott ezen döntések mindig – legalább – a felhasználók szólásszabadságának korlátozásával járnak. A másik legfontosabb különbség a meghatározások között, hogy míg a DSA alapján a tartalommoderálás a moderálással érintett tartalom felderítésére vonatkozó eljárást is a moderálás részének tekinti, addig a felderítés aligha tekinthető moderálásnak abban az értelemben, hogy a tartalommal kapcsolatos tényleges beavatkozás hiányában nem érint alapvető jogokat. A jelen tanulmány céljából, annak alapjogi szemléletére tekintettel a szűkebb, DSA-tól eltérő, beavatkozás nélküli felderítés aktusát nem tekintjük moderálásnak.

A platformok nagyrészt automatizált döntéshozatalán alapuló moderálási döntéseinek ellensúlyozására, illetve az alapvető jogok korlátozásának garanciákkal történő védelme[50] érdekében a DSA alapvetően öt eszközt tartalmaz: (i) tartalommoderálási tájékoztató – eljárási szabályzat; (ii) átláthatósági jelentések; (iii) indokolási kötelezettség; (iv) belső panaszkezelési rendszer és személyzeti felügyelet és (v) peren kívüli vitarendezés.

Ad (i) A 14. cikk kötelezi a közvetítő szolgáltatókat arra, hogy szerződéses feltételeik könnyen hozzáférhető és géppel olvasható formátumban olyan világos, egyszerű, érthető és felhasználóbarát tájékoztatást tartalmazzanak a tartalommoderálásra vonatkozóan, amely magában foglalja valamennyi alkalmazott szabályra, eljárásra, intézkedésre és eszközre – beleértve az algoritmikus döntéshozatalt és az emberi felülvizsgá-

47 Koltay András: A social media platformok jogi státusa a szólásszabadság nézőpontjából. *In Medias Res*, 2019/1. 42.

48 Uo. 47.

49 DSA 3. cikk t).

50 DSA 9. preambulumbekezdés.

latot –, valamint a belső panaszkezelési rendszerük eljárási szabályzatára vonatkozó információkat.[51]

Ad (ii) A mesterséges intelligencia alaptulajdonságai között említendő az összetettség, az adatoktól való függés,[52] az eredendő átláthatatlanság,[53] a kiszámíthatatlanság és a kiismerhetetlenség,[54] amelyekből logikusan adódik, hogy ezek ellensúlyozására[55] az MI szabályozásának általános kialakítása kapcsán az egyik fő erkölcsi eredetű, illetve alapjogi célkitűzés a magyarázhatóság, vagyis a működés átláthatósága.[56] A mesterséges intelligencia átláthatatlansága továbbá alkalmas arra, hogy például közösségi média felhasználók alapvető eljárásjogi jogainak érvényesülését akadályozza, így különösen a hatékony jogorvoslathoz, a tisztességes eljáráshoz való jog érvényesülését.[57]

Ugyanakkor az MI további alapjellemzője az autonóm magatartás is,[58] amely miatt például az átláthatóság biztosítása komoly kihívást jelent különösen gépi tanulásra tekintettel.[59] Ezen problémát erősíti az ún. „feketedoboz jelenségnek" *(black box effect)* nevezett előfeltevés is, amely alapvetően az MI rendszerek azon jellemzőjére utal, hogy az autonóm mesterséges intelligencia rendszerek emberek (felhasználók) által megismerhetetlenül, értelmezhetetlenül működnek.[60] Ezen gondolatmeneten továbbhaladva azzal a paradox helyzettel találhatjuk szembe magunkat (információ-túlterhelés vagy átláthatósági paradoxon[61]), hogy teljes transzparencia, azaz a pontos programozási kód vagy algoritmus nyilvánosságra hozatala sem képes a transzparenciával elérni kívánt cél megvalósítására, hiszen a felhasználók megfelelő ismereteik hiányában nem képesek, vagy pedig csak aránytalan idő- és energiaráfordítással képesek megérteni a rendszerek működését, amely így pontosan az ellenkező hatást éri el: inkább félrevezeti a felhasználókat, mint felvilágosítja őket a rendszer megbízhatóságáról. Továbbá a

---

51 Martin Husovec: Will the DSA Work? In: Joris van Hoboken et al. (szerk.): *Putting the DSA into Practice. Enforcement, Access to Justice and Global Implications*. Berlin, Verfassungsblog gGmbH, 2023. 26. https://doi.org/10.17176/20230208-093135-0

52 European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (a továbbiakban: EC Proposal) 3.5.

53 Papadouli Vasiliki: Transparency in Artificial Intelligence: A Legal Perspective. *Journal of Ethics and Legal Technologies,* Vol. 4., No. 1. (2022) 25.

54 Tóth András: A mesterséges intelligencia szabályozásának paradoxonja és egyes jogi vonatkozásainak alapvető kérdései. *Infokommunikáció és Jog,* 2019/2. 3.

55 Vasiliki i. m. 25.

56 Mezei Kitti: A mesterséges intelligencia jogi szabályozásának aktuális kérdései az Európai Unióban. *In Medias Res,* 2023/1. 55. https://doi.org/10.59851/imr.12.1.4

57 Uo. 62.

58 EC Proposal 3.5.

59 Kovács Zoltán Balázs: Nincs új a nap alatt, avagy figyelemmel az eddigi Európai Uniós hatósági gyakorlatra is, választ ad a GDPR a mesterséges intelligencia használata kapcsán felvetődő adatvédelmi kérdésekre? *Szecskay Ügyvédi Iroda,* 2024. június 18., 8. https://tinyurl.com/3vtuzs3j

60 Alexander J. Wulf – Ognyan Seizov: Artificial Intelligence and Transparency: A Blueprint for Improving the Regulation of AI Applications in the EU. *European Business Law Review,* Vol. 31., No. 4. (2020) 619. https://doi.org/10.54648/EULR2020024

61 Vasiliki i. m. 32.

transzparencia hétköznapi értelemben azt is jelenthetné, hogy a mesterséges intelligenciára vonatkozó tudásanyag szisztematikusan átadásra kerülne egyik érdekelt féltől a másiknak,[62] amely azonban azt eredményezheti, hogy az MI rendszerekre vonatkozó nem megfelelő mértékű és módon történő átláthatósági követelmények előírása aránytalan sérelmet jelentene a nem nyilvános know-how és üzleti információk (üzleti titkok) felfedésére kötelezéssel.[63]

A DSA az automatizált döntéshozatalán alapuló moderálási döntéseinek ellensúlyozására átláthatósági jelentési kötelezettségeket ír elő a szolgáltatókra vonatkozóan.[64] Ezen kötelezettség három szinten kerül meghatározásra kumulatív jelleggel,[65] egyrészt valamennyi közvetítő szolgáltatóra vonatkozó minimális tartalmi elemek keretében,[66] másrészt az online platformot üzemeltető szolgáltatókra vonatkozóan,[67] és harmadrészt pedig az online óriásplatformot vagy nagyon népszerű online keresőprogramot üzemeltető szolgáltatókra irányadóan.[68]

Az óriásplatformnak minősülő, legnépszerűbb közösségi szolgáltatók az átláthatósági jelentésekben az automatizált döntéshozatallal kapcsolatosan az alábbi információkat kötelesek feltüntetni:

a) érdemi és érthető tájékoztatást a szolgáltató saját kezdeményezésére végzett tartalommoderálásról, beleértve az automatizált eszközök használatát is;[69]
b) arra vonatkozó információt, hogy automatizált eszközöket alkalmaz tartalommoderálás céljából, beleértve a minőségi leírást, a pontos célok meghatározását, az e célok eléréséhez használt automatizált eszközök pontosságára és lehetséges hibaarányára vonatkozó mutatókat, valamint az alkalmazott biztosítékokat;[70]
c) a (b) pontban írt automatizált eszközökre vonatkozóan pontossági mutatókra és kapcsolódó információkra kiterjedő tájékoztatást.[71]

Ad (iii) A DSA megalkotásának egyik fő célkitűzése a hatékony jogorvoslati mechanizmusok hozzáférhetőségének javítása volt.[72] Azért, hogy a felhasználók hatékonyan tudjanak élni jogorvoslati jogukkal, valamennyi tárhelyszolgáltató köteles a döntését (nem csak a tartalmak törlése, hanem például a demonetizáció vagy a tartalom hátra sorolása esetén is) megfelelően megindokolni, ha moderálást végez, ugyanis az indokolásnak alapvető jelentősége van a jogorvoslati jog gyakorlása során. A közösségimédia-

---

62 Reid Blackman – Beena Ammanath: Building Transparency into AI Projects. *Harvard Business Review,* 2022. június 20. https://tinyurl.com/27xvyrst

63 EC Proposal 3.5. pont.

64 DSA 65. preambulumbekezdés.

65 Ződi Zsolt: Átláthatósági és indokolási követelmények az európai platformjogban. *In Medias Res*, 2023/2. 17. https://doi.org/10.59851/imr.12.2.1

66 DSA 15. cikk.

67 DSA 24. cikk.

68 DSA 42. cikk.

69 DSA 15. cikk (1) c).

70 DSA 15. cikk (1) e).

71 DSA 42. cikk (2) c).

72 DSA 9. preambulumbekezdés.

szolgáltatók indokolási kötelezettsége mindazokra a korlátozásokra fennáll, amelyeket a felhasználó által megosztott tartalom (információ) jogellenessége vagy a szerződési feltételekkel való összeegyeztethetetlensége miatt alkalmaznak, és fontos eleme, hogy a közvetítő szolgáltató a bejelentést követő passzivitását is köteles indokolni. A DSA a jogi döntésekre előírt felépítést, illetve logikát követi azzal a kivétellel, hogy az alkalmazott jogkövetkezményt is az indokolás tartalmazza a rendelkező rész helyett. Ezen túl kötelező elem a tényállás megállapítása, a jogalap vagy szerződéses feltétel megjelölése és külön pontban foglalnak helyet a döntés meghozatala során használt automatizált (mesterséges intelligencia vezérelt) eszközök alkalmazására vonatkozó információk. Ugyancsak a jogi döntések mintájára az indokolásnak tartalmaznia kell egy jogorvoslati klauzulát.

Ad (iv) Míg a DSA 17. cikkében előírt indokolási kötelezettség nagyrészt a szolgáltatók aktivitása esetén irányadó, a DSA 20. cikkében foglalt, legalább hat hónapig biztosítandó, elektronikus és ingyenes hatékony panaszmechanizmus[73] a döntések szélesebb körére vonatkozik, és abban az esetben is lehetőséget nyújt a panasz benyújtására, ha a 16. cikk szerinti bejelentés alapján a platform passzív marad.[74] A panaszeljárás szükségképpen személyzeti közrehatást igényel, és a szolgáltató indokolatlan késedelem nélkül köteles tájékoztatni a panaszost indokolt döntéséről és a peren kívüli vitarendezés lehetőségeiről vagy egyéb jogorvoslati lehetőségről. Az „indokolatlan késedelem nélkül" kitétel azonban legalább annyira tágan megfogalmazott követelmény, mint az Eker. irányelvben foglalt haladéktalan eltávolítási kötelezettség, és konkretizálás hiányában várhatóan számos bizonytalanság forrása lesz.[75]

A 17. cikkben írtak alapján jogosan merül fel a kérdés, hogy mennyiben tud érdemben panaszt tenni az olyan bejelentést tevő, akinek a bejelentése alapján a szolgáltató nem tesz intézkedést, azonban a 17. cikk alapján indokolási kötelezettsége sem volt a döntés kapcsán. A korlátozás és a közvetítő szolgáltató passzivitása[76] esetére előírt eltérő indokolási kötelezettség azonban azzal az aszimmetrikus következménnyel járhat, hogy a felhasználó nem ugyanolyan esélyekkel tud jogorvoslattal élni a szolgáltató passzivitása (tartózkodása, nem tevése) esetén, mint az a felhasználó, akinek tartalmát a szolgáltató indokolt döntéssel távolította el.[77] A kötelezettségek összességét megvizsgálva azonban láthatjuk, hogy a szolgáltató köteles a körülményekre és a tényállásra kiterjedő indokolást nyújtani a belső panaszkezelési mechanizmus keretében eldöntött panasz kapcsán, így komplementálva a teljes indokolási kötelezettség hiányát passzivitás esetén (hiszen a peren kívüli vitarendezés igénybevételének előfeltétele a panasz-

[73] DSA 20. cikk (1).

[74] DSA 17. cikk.

[75] Részletesen ld. Gosztonyi Gergely: A közösségi média felelősségi kérdéseinek korai szabályozása az Amerikai Egyesült Államokban és az Európai Unióban. In: Mezey Barna (szerk.): *Kölcsönhatások. Európa és Magyarország a jogtörténelem sodrásában*. Budapest, Gondolat, 2021. 111–119.

[76] Pietro Ortolani: If You Build It, They Will Come. In: Joris van Hoboken et al. (szerk.): *Putting the DSA into Practice. Enforcement, Access to Justice and Global Implications*. Berlin, Verfassungsblog gGmbH, 2023. 151–166. http://www.doi.org/10.17176/20230208-093135-0

[77] DSA 16. cikk (5).

mechanizmus kimerítése),[78] ezáltal lehetővé téve a felhasználónak, hogy a 21. cikk szerinti peren kívüli vitarendezési fórum előtt egy indokolt döntést tudjon vitatni.[79]

Ad (v) A peren kívüli vitarendezésnek azért is van kiemelkedő jelentősége, mert korábban – ide nem értve a Facebook (Meta) által felállított Ellenőrző Bizottság közel négy évnyi tevékenységét – a felhasználók és a közösségi média platformok közötti vitarendezés egyetlen módja az állami (és esetlegesen választott-) bíróságon keresztül történt. 2024. február 17-től, a DSA teljeskörű alkalmazását követően a felhasználók peren kívüli vitarendezésre jogosultak, amelyet a rendeleti forma miatt minden tagállamnak biztosítania kell úgy, hogy a költségek szignifikáns részét a közösségi médiaszolgáltatók kötelesek viselni.[80] A rendelet által felvázolt peren kívüli vitarendezés mechanizmusa minden mikro- és kisvállalkozásnál nagyobb online platformot üzemeltető szolgáltatóra vonatkozik,[81] aki/amely az EU-ban található, vagy ott van a letelepedési helye, függetlenül az említett közvetítő szolgáltatást biztosító szolgáltatók letelepedési helyétől.[82]

Ugyan a DSA 21. cikkében foglalt testület előtti eljárás eredményeként meghozott döntések – az alávetés kivételével – kötelező erővel nem bírnak, lehetővé teszik, hogy releváns szakértelemmel rendelkező tagokból álló bizottság idő- és költséghatékony módon, a bírósági szervezetrendszeren kívül dönthessen olyan vitás kérdésekben, amelyeket az online platform felhasználóinak a szolgáltatóval közvetlenül nem sikerül rendezniük.[83] Ezen eljárások a bírósághoz fordulás jogának sérelme nélkül vehetők igénybe, de mindenképpen azzal szembeni előnyük, hogy egy szakértői testület hoz döntést, míg a bíróság legfeljebb csak bizonyos szakkérdésben támaszkodhat szakértői véleményekre.[84]

3.2. Az MI rendelet és a tartalommoderálás, illetve az átláthatóság

Az MI rendelet jelentős változásokat hoz az online platformok mesterséges intelligencia tartalommoderálásban való felhasználásának módjában, ami lehetőségeket és kihívásokat egyaránt rejt magában. Az MI rendelet egy kockázatalapú megközelítést alkalmaz, és az AI-rendszereket a gyakorlatban négy kategóriába sorolja: elfogadha-

---

78 Lendvai Gergely Ferenc: A Facebook Ellenőrző Bizottság működése és bíráskodása a gyűlöletbeszéd kontextusában. *In Medias Res,* 2024/1. 195–221. https://doi.org/10.59851/imr.13.1.11

79 Husovec i. m. 24.

80 Hannah Ruschemeier et al.: Brave New World: Out-Of-Court Dispute Settlement Bodies and the Struggle to Adjudicate Platforms in Europe. *Verfassungsblog,* 2024. szeptember 10. https://tinyurl.com/yt9zn7pb http://www.doi.org/10.59704/46b8611eb2d96a84

81 Gyetván Dorina: A Digital Services Act és a Facebook Oversight Board szerepe a jogorvoslat biztosításában a közösségi média vonatkozásában. *In Medias Res,* 2023/2. 190–208. https://doi.org/10.59851/imr.12.2.9

82 DSA 21. cikk.

83 DSA 21. cikk (2).

84 Gyetván Dorina – Gosztonyi Gergely: Garanciális elemek az online platformok véleménynyilvánítást korlátozó moderálási gyakorlatával szemben a digitális szolgáltatásokról szóló rendelet (DSA) előtt és után. *Magyar Jog,* 2025/4. 226–236. https://doi.org/10.59851/mj.72.04.3

tatlan, nagy, rendszerszintű és minimális kockázatú rendszerek.[85] Az online platformok által használt tartalommoderációs eszközök általában a magas vagy a korlátozott kockázatú rendszerek kategóriájába tartoznak, attól függően, hogy milyen hatással lehetnek az alapvető jogokra, például a véleménynyilvánítás szabadságára és a megkülönböztetésmentességre.

Az MI rendelet értelmében tiltott az olyan mesterséges intelligencia rendszerek használata, amelyek elfogadhatatlan kockázatot jelentenek az alapvető jogokra nézve. Ide tartoznak az olyan mesterséges intelligenciával működő rendszerek, amelyek manipulálják az emberi viselkedést, vagy olyan módon használják ki a kiszolgáltatott személyeket, ami kárt okozhat nekik.[86] A tartalommoderálással összefüggésben ebbe a kategóriába tartoznának azok az AI-rendszerek, amelyek szándékosan vagy szisztematikusan cenzúrázzák a törvényes tartalmakat a közvélemény manipulálása érdekében.[87]

A nagy kockázatú mesterséges intelligencia rendszerekre a legszigorúbb szabályok vonatkoznak.[88] Az online óriásplatformok, például a Facebook, a YouTube és az X/Twitter által használt tartalommoderációs rendszerek a felhasználók alapvető jogaira gyakorolt potenciális hatásuk miatt a magas kockázatúak közé sorolhatók. Ezek a platformok gyakran használnak mesterséges intelligenciát a közösségi normákat sértő tartalmak, például a gyűlöletbeszéd, a félretájékoztatás és a szélsőséges tartalmak automatikus felismerésére és eltávolítására. Tekintettel e funkciók érzékenységére, az EU különleges kötelezettségeket ró a magas kockázatú mesterséges intelligencia rendszerekre, különösen az átláthatóság, a felügyelet és a pontosság tekintetében.[89]

A felhasználók jogait kevésbé befolyásoló tartalommoderációs eszközök a rendszerszintű és a minimális kockázatú rendszerek közé tartoznak.[90] Ezek közé tartozhatnak olyan alapvető automatikus szűrőrendszerek, amelyek például blokkolják a spameket vagy észlelnek bizonyos szerzői jogi jogsértéseket. A korlátozott kockázatú rendszerek esetében a követelmények enyhébbek, de továbbra is tartalmaznak átláthatósági kötelezettségeket. A minimális kockázatú mesterséges intelligenciára, például az ajánlórendszerekben használt általános célú algoritmusokra vagy a korlátozott moderálási funkcióval rendelkező nyelvi szűrőkre még kevesebb szabályozási kötelezettség vonatkozik.[91]

---

85 Martin Ebers: Truly Risk-Based Regulation of Artificial Intelligence. How to Implement the EU's AI Act. *SSRN,* 2024. június 19., 7–11. http://dx.doi.org/10.2139/ssrn.4870387

86 MI rendelet 5. cikk.

87 Michael Veale – Frederik Zuiderveen Borgesius: Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International,* Vol. 22., No. 4. (2021) 97–112. https://doi.org/10.9785/cri-2021-220402

88 MI rendelet 6–27. cikkek.

89 Gerhard Wagner: Liability for Artificial Intelligence. A Proposal of the European Parliament. *Working Paper des Forschungsinstituts für Recht und digitale Transformation,* 2021/9. 1–34. https://doi.org/10.2139/ssrn.3886294

90 MI rendelet 51–55. cikkek.

91 Luciano Floridi: The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology,* Vol. 34., No. 1. (2021) 215–222. https://doi.org/10.1007/s13347-021-00460-9

Az MI rendelet több cikke közvetlenül foglalkozik a mesterséges intelligenciát tartalom moderálására használó platformok felelősségével, az átláthatóságra, az adatkezelésre, az elszámoltathatóságra és az emberi felügyeletre összpontosítva.[92]

A tartalom moderálásához használt mesterséges intelligencia rendszerek tisztességes és elfogulatlan működésének biztosításához elengedhetetlen a megfelelő adatkezelés.[93] Az MI rendelet 10. cikke előírja, hogy a magas kockázatú mesterséges intelligencia rendszerek képzéséhez felhasznált adatoknak jó minőségűnek, elfogultságtól mentesnek és az összes releváns csoportot reprezentálónak kell lenniük. Ez különösen fontos a tartalommoderálásban, ahol az elfogult képzési adatok bizonyos csoportokkal szembeni tisztességtelen bánásmódot vagy a marginalizált közösségek túlságosan szigorú ellenőrzését eredményezhetik.[94] Ha például a gyűlöletbeszéd megjelölésére használt mesterséges intelligencia rendszert olyan adatokon képzik ki, amelyekben túlnyomórészt bizonyos dialektusok vagy nyelvek szerepelnek, akkor az aránytalanul nagy mértékben megjelölheti az ezekből a csoportokból származó tartalmakat, ami aggályokat vethet fel a diszkriminációval és a szólásszabadsággal kapcsolatban.

Az automatizált döntésekre való túlzott hagyatkozás megelőzése érdekében az MI rendelet 14. cikke előírja, hogy a nagy kockázatot jelentő mesterséges intelligencia rendszereket emberi felügyeletnek kell alávetni.[95] A tartalom moderálásával összefüggésben ez azt jelenti, hogy bár a mesterséges intelligencia automatizálhatja a káros tartalmak felismerését, a végső döntések meghozatalában – különösen a felhasználók alapvető jogait érintő döntésekben – emberi moderátoroknak kell részt venniük. Ez az ember által alkalmazott megközelítés elengedhetetlen a mesterséges intelligencia korlátainak kezeléséhez, például ahhoz, hogy nem képes teljes mértékben felfogni vagy helyén kezelni a kontextust. A tartalommoderálás esetében ez azt jelenti, hogy az AI-eszközök megbízhatóan felismerik a káros tartalmakat, miközben minimalizálják a hamis pozitív eseteket, amikor a jogszerű tartalmakat tévesen jelölik vagy távolítják el, és a hamis negatív eseteket, amikor a káros tartalmakat nem veszik észre.[96] Az emberi felügyeletre vonatkozó rendelkezések megerősítik az automatizált döntések és az emberi ítélőképesség egyensúlyának fontosságát. Míg a mesterséges intelligencia hatékonyan képes nagy mennyiségű tartalom moderálására, az emberi moderátorok továbbra is nélkülözhetetlenek az összetett, kontextuális megértést igénylő ügyek kezeléséhez.

---

92 Solymár Károly Balázs: A fogalomhasználat egyes kérdései a mesterséges intelligenciáról szóló európai rendelettervezetben. In: Glavanits Judit (szerk.): *Fogyasztóbarát mesterséges intelligencia – a velünk élő AI egyes aktuális kérdései*. Győr, Universitas-Győr Nonprofit Kft., 2023. 9–27.

93 Theodore S. Boone: The challenge of defining artificial intelligence in the EU AI Act. *Journal of Data Protection and Privacy,* Vol. 6., No. 2. (2023) 180–195. https://doi.org/10.69554/QHAY8067

94 Reuben Binns: Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of Machine Learning Research,* Vol. 10., No. 81. (2018) 149–159.

95 Tóth András: Az Európai Unió Mesterséges Intelligencia Törvényéről. *Gazdaság és Jog,* 2024/5–6. 3–11.

96 Tarleton Gillespie: *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven, Yale University Press, 2018. 99. https://doi.org/10.12987/9780300235029

Az MI rendelet bizonyos, tartalommoderálási kérdésekkel is összefüggésbe hozhatóan, a kockázatok csökkentése érdekében viszonylag hangsúlyosan tartalmaz átláthatósági követelményeket, meghatározásokat a DSA-hoz hasonlóan. A fogalom tisztázása érdekében például a 27. preambulumbekezdés leszögezi, hogy az MI rendelet keretei között az átláthatóság azt jelenti, hogy az MI-rendszereket olyan módon kell fejleszteni és használni, amely lehetővé teszi a megfelelő nyomonkövethetőséget és megmagyarázhatóságot, miközben tudatosítja az emberekben, hogy MI-rendszerrel kommunikálnak vagy lépnek kapcsolatba, valamint megfelelően tájékoztatja az alkalmazókat az MI-rendszer képességeiről és korlátairól, az érintett személyeket pedig jogaikról.[97]

Erre építve az MI rendelet 50. cikke a különböző érdekeltek jogainak és érdekeinek egyensúlyban tartására, az információs aszimmetria mérséklésére törekedve tartalmaz egyes olyan MI-rendszerekre vonatkozóan átláthatósági kötelezettségeket, amelyek (a) emberekkel érintkeznek, (b) az érzelmek észlelésére vagy a biometrikus adatokon alapuló (szociális) kategóriákkal való kapcsolat meghatározására szolgálnak, vagy (c) tartalmakat hoznak létre vagy manipulálnak.[98] Ha a emberek MI-rendszerrel érintkeznek, vagy érzéseiket vagy jellemzőiket automatizált eszközökkel ismerik fel, erről őket tájékoztatni kell. Ha az MI-rendszert olyan képek, audio- vagy videotartalmak létrehozására vagy manipulálására használják, amelyek érzékelhetően hasonlítanak hiteles tartalmakra, kötelezővé kell tenni annak közlését, hogy a tartalom automatizált módon jött létre, eltekintve a jogszerű célokat szolgáló kivételektől. Ez lehetővé teszi a személyek számára, hogy megalapozott döntéseket hozzanak, vagy visszalépjenek adott helyzetből.[99]

4. A mesterséges intelligencia a tartalommoderálás gyakorlatában

A mesterséges intelligencia központi eszközzé vált az online platformok tartalommoderációjában, ahol a felhasználók által generált hatalmas mennyiségű tartalom kezelése kritikus kihívást jelent. Az olyan platformok, mint a Facebook, az X/Twitter, a YouTube és az Instagram naponta hatalmas mennyiségű posztot, képet, videót és kommentet dolgoznak fel, és az AI hatékony módot kínál a közösségi irányelveket sértő tartalmak ellenőrzésének, megjelölésének és eltávolításának automatizálására. Bár a mesterséges intelligencia nagymértékben javíthatja a tartalom moderálását, nem mentes a kihívásoktól. Alkalmazása sikeres és hibás tartalomosztályozást egyaránt eredményezett, ami számtalan vitához vezetett a pontosság, a tisztességesség és az etikai vonatkozások körül. Jelen fejezetben vizsgáljuk a tartalom moderálásának hatékonyságát, korlátait és azokat a konkrét eseteket, amikor a mesterséges intelligencia jól vagy rosszul teljesített.

A mesterséges intelligencia a tartalom moderálásában általában a gépi tanulási algoritmusok, a természetes nyelvi feldolgozás és a számítógépes látás kombinációjával működik. Ezek a technológiák teszik lehetővé a mesterséges intelligencia számára,

97 MI rendelet 27. preambulumbekezdés.

98 MI rendelet 50. cikk.

99 EC Proposal 5.2.4.

hogy nagy mennyiségű adatot gyorsan megvizsgáljon és elemezzen, és olyan mintákat vagy jeleket azonosítson, amelyek a közösségi irányelvek megsértésére utalhatnak.

Az algoritmusokat nagy adathalmazokon tanítják be, amelyek az elfogadható és elfogadhatatlan tartalom címkézett példáit tartalmazzák. A mesterséges intelligencia ezután e tanult minták alapján megjósolja, hogy az új tartalom megfelel-e a közösségi irányelveknek.[100] Az ilyen modellek folyamatosan fejlődnek, ahogy új adatokkal találkoznak, lehetővé téve a tartalomfelismerés dinamikus fejlesztését. A természetes nyelvi feldolgozás (*Natural Language Processing*, a továbbiakban: NLP) elengedhetetlen a szövegalapú tartalom elemzéséhez és a nem megfelelő kifejezések felismeréséhez. Az NLP modellek néha képesek felismerni egy szó vagy kifejezés kontextusát, megkülönböztetve a jóindulatú és a káros felhasználásokat.[101] Az X/Twitter például NLP-t használ a sértő nyelvezetet vagy gyűlöletbeszédet tartalmazó tweetek megjelölésére. A képek és videók moderálásához a számítógépes látás algoritmusai elemzik a vizuális tartalmat, hogy felismerjék a nem megfelelő képi anyagokat, például a meztelenséget, a grafikus erőszakot vagy a terrorista propagandát.[102]

A mesterséges intelligencia számos esetben nagyon hatékony a káros tartalmak azonosításában. A hatalmas mennyiségű tartalom gyors feldolgozásának képessége elengedhetetlen a több milliárd felhasználóval rendelkező platformok számára. A Facebook például sajtóhírek szerint jelentős összegeket fektetett be a mesterséges intelligenciába a terrorizmussal kapcsolatos tartalmak felderítése és eltávolítása érdekében. A vállalat 2024-ben arról számolt be, hogy a terrorizmussal kapcsolatos posztok több mint 99%-át az AI megjelölte, mielőtt még azokat a felhasználók jelentették volna.[103] A COVID-19 világjárvány idején a YouTube mesterséges intelligenciát használt az egészségügyi félretájékoztatás terjedése elleni küzdelemhez, és a mesterséges intelligencia segítségével azonosították az irányelvek megsértése miatt megjelölt videók 94%-át.[104] A Twitter mesterséges intelligenciája 2024-ben a jogsértő felhasználói fiókok több mint 99%-át proaktívan kezelte anélkül, hogy felhasználói jelentésekre támaszkodott volna.[105]

E sikerek ellenére a mesterséges intelligencia alapú tartalommoderáció még mindig jelentős korlátokkal küzd, ami nagy visszhangot kiváltó kudarcokhoz vezetett. Ezek gyakran abból fakadnak, hogy a mesterséges intelligencia nem képes megérteni a

---

100 Robert Gorwa – Reuben Binns – Christian Katzenbach: Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society,* Vol. 7., No. 1. (2020) https://doi.org/10.1177/2053951719897945

101 Anna Schmidt – Michael Wiegand: A Survey on Hate Speech Detection using Natural Language Processing. In: Lun-Wei Ku – Cheng-Te Li (szerk.): *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Valencia, Association for Computational Linguistics, 2017. 1–10. https://doi.org/10.18653/v1/W17-1101

102 Theo Araujo et al.: In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society,* Vol. 35., No. 3. (2020) 611–623. https://doi.org/10.1007/s00146-019-00931-w

103 Facebook: *Community Standards Enforcement Report. Q2 2024 report.* https://tinyurl.com/22k4j7kp

104 Lambodara Parabhoi et al.: YouTube as a source of information during the Covid-19 pandemic: a content analysis of YouTube videos published during January to March 2020. *BMC Medical Informatics and Decision Making,* Vol. 21., No. 1. (2021) 255–265. https://doi.org/10.1186/s12911-021-01613-8

105 X: *Global Transparency Report, H1 2024*. https://tinyurl.com/2p86w9bx

tartalom teljes kontextusát, a kulturális árnyalatokat és a nyelv bonyolultságát. Bár e rendszerek folyamatosan fejlődnek, a hibás tartalommoderáció számos reputációs problémát tud okozni a vállalatoknak is. 2016-ban a Facebook mesterséges intelligenciája helytelenül eltávolította a vietnami háborúban készült ikonikus Napalm Girl fotót, arra hivatkozva, hogy meztelensége sérti a tartalmi irányelveket.[106] Az MI nem ismerte fel a kép történelmi jelentőségét, és kizárólag a meztelenség alapján nem megfelelő tartalomként kezelte. 2019-ben az égő párizsi Notre Dame katedrális képeit távolította el az MI, mivel – helytelenül –a 2001-es amerikai terrortámadás képeinek értelmezte azokat.[107] A 2020-as COVID-19 világjárvány idején a YouTube szintén nagymértékben támaszkodott a mesterséges intelligenciára a tartalom moderálásában, mivel az emberi moderátorok otthonról dolgoztak. Ez a hamis pozitív eredmények jelentős növekedéséhez vezetett, amikor a COVID-19-ről szóló oktatási és hírekkel kapcsolatos tartalmakat tévesen félretájékoztatásként távolították el. De a szatirikus tartalom is gyakran kihívást jelent a mesterséges intelligencia számára, ugyanis a szatírát – annak a normál beszédtől eltérő kontextusáról és hangneméről szinte tudomást sem véve – gyakran azonosítja a mesterséges intelligencia jogsértő tartalomként.[108]

A mesterséges intelligenciának a tartalom moderálásával kapcsolatos nehézségei nagyrészt abból erednek, hogy a betáplált adatmodellek hatalmas mennyisége ellenére sem képes teljes mértékben megérteni a kontextust, az eltérő kultúrákat és a nyelvi különbözőségeket. Így például fontos lenne annak ismerete, hogy a nyelv folyamatosan fejlődik, és gyorsan jelennek meg új szleng- és köznyelvi kifejezések, különösen a közösségi médiában. Az elavult adathalmazokon kiképzett mesterséges intelligencia-modellek nem feltétlenül képesek megragadni az új kifejezéseket, ami pontatlan moderáláshoz vezethet. Ráadásul a különböző kultúrákban a nyelvezet jelentősen eltér, ami megnehezíti, hogy a globális platformok minden országban egyformán és egységesen alkalmazható moderációs rendszereket fejlesszenek ki. Nem szabad elfelejteni, hogy a mesterséges intelligencia-modellek csak annyira jók, mint az adatok, amelyeken kiképzik őket. Ha az adatok elfogultak vagy hiányosak, a végeredmény is elfogult eredményeket fog produkálni.

5. Összefog(lal)ás? A mesterséges intelligencia és az emberi moderálás kapcsolatának jövője

A mesterséges intelligencia korlátai miatt sok platform hibrid megközelítést alkalmaz a tartalom moderálására, amely a mesterséges intelligencia sebességét emberi felügyelettel kombinálja. Az emberi moderátorok felülvizsgálják a megjelölt tartalmakat, hogy biztosítsák a kontextuális és nyelvi árnyalatok, illetve a kulturális különbségek figyelembevételét. Ez az együttműködő megközelítés lehetővé teszi a platformok számára,

106 Gillespie i. m. 1–4.

107 Tuesday briefing: YouTube's fake news detection tool flagged the Notre Dame fire with 9/11 facts. *Wired,* 2019. április 16., https://tinyurl.com/5547uudk

108 Tom Lymn – Jessica Bancroft: The use of algorithms in the content moderation process. *gov.uk,* 2021. augusztus 5. https://tinyurl.com/hryn9uy8

hogy egyensúlyba hozzák a mesterséges intelligencia erősségeit az árnyalt döntéshozatal emberi képességével. A mesterséges intelligencia fejlődésével a tartalommoderációs rendszerek valószínűleg egyre kifinomultabbá válnak. A mélytanulás, a kontextuális elemzés és a mesterséges intelligencia kulturális különbségek megértésére való képességének javulása növelheti a moderációs rendszerek pontosságát.

Az EU mesterséges intelligencia rendelete és a DSA fontos szerepet fognak játszani az online platformokon a mesterséges intelligencia által vezérelt tartalommoderáció jövőjének alakításában. A rendeletek szigorú követelményeket írnak elő a mesterséges intelligenciával működő rendszerekkel szemben, és azt kívánják biztosítani, hogy a tartalommoderációs eszközök átláthatóak, tisztességesek és elszámoltathatók legyenek. Ezzel lehetőséget kínálnak arra, hogy a tartalommoderálás minőségének javítása és az alapvető jogok védelme a digitális korban is biztosított legyen.

Összefoglalva, bár a mesterséges intelligencia létfontosságú szerepet játszik az online platformok tartalmi moderálásában, annak használata és használhatósága nem korlátlan.[109] A mesterséges intelligencia és az emberi moderálás közötti egyensúly – ahogy az online platformok alkalmazkodnak majd az új szabályozásokhoz – továbbra is fejlődni fog, különösen azáltal, hogy a platformok is rá lesznek kényszerítve arra, hogy emberi közreműködésen keresztül enyhítsék a felmerülő technológia-indukált problémákat. Az emberi felügyelet a közeljövőben továbbra is kulcsfontosságú marad a mesterséges intelligencia által még meg nem érthető finomságok és összetettségek kezeléséhez.

[109] Content moderation in a new era for AI and automation. *Oversight Board,* 2024. szeptember 17. https://tinyurl.com/yc5b4nbs