

GOSZTONYI GERGELY

A tartalomszabályozás emberi és technikai vonatkozásai

Human and Technical Aspects of Content Regulation

Abstract: Facebook had 2.8 billion users in 2020, and the company generates four petabytes of data a day. In addition to the legal regulation, one pays less attention to the issues of human moderators and artificial intelligence although the classification of content is done by them through the combined use of artificial intelligence and human intervention. Together they are the first line of defence in the war on unwanted content. The solution is probably not a decision between human moderation and moderation by artificial intelligence but for using them jointly in the future.

Keywords: Internet, content regulation, artificial intelligence, moderators, moderation

Összefoglaló: A Facebooknak 2,8 milliárd felhasználója volt 2020-ban, és a cég naponta négy petabájt adatot generál. A jogi szabályozás mellett – azt kiegészítve – kevesebb figyelmet fordítunk az emberi moderátorok és a mesterséges intelligencia kérdéseire, pedig a tartalmak osztályozását a mesterséges intelligencia és emberi beavatkozás együttes használatával ők végzik. Ők együtt jelentik az első védvonalat a kérértlen tartalom elleni háborúban. A megoldás valószínűleg nem az emberi moderálás vagy a mesterséges intelligencia általi moderálás közötti döntés, hanem a kettő együttes használatát jelenti majd a jövőben.

Kulcsszavak: internet, tartalomszabályozás, mesterséges intelligencia, moderátorok, moderálás

1. Bevezető

Peter Paul Rubens világhírű flamand festő 1612 és 1614 között festette meg a *Levétel a keresztről* című festményét, 1614–15 körül pedig a *Vénusz a tükör előtt* címet viselőt. Valószínűleg nem csupán a közösségi média akkori ismeretlensége miatt sem gondolta soha, hogy festménye – pontosabban a festménye körüli események – néhány évszázaddal később közneveltség tárgyává tesznek egy vállalatot. 2018-ban ugyanis az állami Flamand Turisztikai Iroda reklámkampányt indított a Facebookon, amelyben felhasználta ezt a két képet is, ám a Facebook mindkét festményt eltávolította, arra

hivatkozva, hogy azok indokolatlan meztelenséget tartalmaznak.¹ Az iroda a cégnek írt levelében így fogalmazott: „*Annak ellenére, hogy titokban nevetünk ezen, az Önök kulturális cenzúrája meglehetősen megnehezíti számunkra az életet.*”² Rubens alkotásához hasonló sorsra jutott korábban Gustave Courbet 1866-os festménye, *A világ eredete*³ (amely kapcsán nyolc évig tartó bírósági eljárás is zajlott Franciaországban⁴) vagy a *Willendorfi Vénusz*ról készült fotó⁵ is, mivel a meztelen testrészek egyértelmű ábrázolását a pornográfia kategóriájába tartozónak minősítették a cég moderálási irányelvei (Community Standards, CS).⁶ De nem csak festmények kerültek hasonló helyzetbe: a dán főváros, Koppenhága ikonikus szobra, a *Kis hableány* is eltávolításra került, mivel „*túl sok meztelen bőrfelületet vagy szexuális felhangot*” tartalmazott a Facebook szerint.⁷ Eltávolításra került – erőszakos tartalom miatt – a jászolban fekvő csecsemő Jézus előtt térdelő Mikulásról készült festmény is.⁸ De akár az is könnyen elképzelhető, hogy valaki egy macskás fotó miatt kerül tiltólistára: egy felhasználónak 2016-ban zárolták a felhasználói fiókját egy öltönyös macska képének megosztása miatt.⁹ Vagy olyan tartalmat töröl le a cég – az amerikai rendőrség kérésére –, amelyben egy fekete nő a fehér rendőrökről készített videofelvételt.¹⁰ A sor még folytatható.

¹ N/A: *Facebook Censors Rubens Paintings For Nudity*. frieze, 2018. július 24., <https://www.frieze.com/article/facebook-censors-rubens-paintings-nudity> (letöltés ideje: 2021. 07. 27.).

² *Uo.* Hozzá kell ugyanakkor tenni, hogy az események kapcsán a flamand múzeumok forgalma szerencsésen növekedett az adott időszakban, és az iroda elkészített egy videót, amelyben kifigurázzák a Facebook szabályzatát: <https://www.youtube.com/watch?v=UZq3cVgU5AI> (letöltés ideje: 2021. 07. 27.).

³ Victoria STAPLEY-BROWN: *Paris court hears arguments in Facebook censorship case centering on Courbet's Origin of the World*. The Art Newspaper, 2018. február 2., <https://www.theartnewspaper.com/news/paris-court-hears-arguments-in-facebook-censorship-case-centering-on-courbet-s-origin-of-the-world> (letöltés ideje: 2021. 07. 27.).

⁴ Az érdekes és a felek megegyezésével záródó eljárásról ld.: Sarah CASCOE: *After an Eight-Year Legal Battle, Facebook Ends Its Dispute With a French School Teacher Who Posted Courbet's 'Origin of the World'*. artnet, 2019. augusztus 5., <https://news.artnet.com/art-world/facebook-courbet-lawsuit-ends-1616752> (letöltés ideje: 2021. 07. 27.).

⁵ Aimee DAWSON: *Facebook censors 30,000 year-old Venus of Willendorf as 'pornographic'*. The Art Newspaper, 2018. február 27., <https://www.theartnewspaper.com/news/facebook-censors-famous-30-000-year-old-nude-statue-as-pornographic> (letöltés ideje: 2021. 07. 27.).

⁶ Facebook Community Standards, III.14. Adult nudity and sexual activity, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity/ (letöltés ideje: 2021. 07. 27.).

⁷ N/A: Denmark: *Facebook blocks Little Mermaid over 'bare skin'*. BBC, 2016. január 4., <https://www.bbc.com/news/blogs-news-from-elsewhere-35221329> (letöltés ideje: 2021. 07. 27.).

⁸ Angelo STAGNARO: *Facebook Censors Image of Santa Kneeling Before Our Lord*. National Catholic Register, 2018. december 26., <https://www.ncregister.com/blog/facebook-censors-image-of-santa-kneeling-before-our-lord> (letöltés ideje: 2021. 07. 27.).

⁹ Michael MOORE: *Facebook sharing this cat photo could get you BANNED*. Daily Express, 2016. október 5., <https://www.express.com.uk/life-style/science-technology/717978/facebook-bans-user-sharing-cat-photo> (letöltés ideje: 2021. 06. 27.).

¹⁰ Timothy KARR: *How Censoring Facebook Affects the Fight for Black Lives*. The Root, 2016. augusztus 29., <https://www.theroot.com/how-censoring-facebook-affects-the-fight-for-black-live-1790856542>

Mindezek miatt – és az amerikai National Coalition Against Censorship¹¹ civil szervezet kampányának következtében – a Facebook egyetértett abban, hogy a CS felülvizsgálatra szorul,¹² ám a folyamat időigényes.¹³ Ahogy a *Daily Mail*-ben megfogalmazták: „Végül is a fenék az egy fenék, a mellbimbó pedig egy mellbimbó. De annak eldöntése nem könnyű még az embereknek sem, hogy egy mellbimbó mikor művészet, mikor pornó vagy mikor tiltakozás. És a mesterséges intelligenciát megtanítani az emberi szexualitásról egy teljesen más kérdés.”¹⁴

De miért érdekes mindez? A Facebooknak 2,8 milliárd felhasználója volt 2020-ban, és a cég „naponta négy petabájt adatot generál – ez millió gigabyte. Ezeket az adatokat az úgynevezett Hive tárolja, amely körülbelül 300 petabájt adatot tartalmaz.”¹⁵ Ezen óriási adathalmaz jelentős része nyilvánvalóan jogszerű tartalom, ám még az a kis rész is hatalmas, aminek nem szabadna felkerülnie. Ezen tartalmak vizsgálatával kapcsolatban azonban az elmúlt időszakban az úgynevezett „privatizált” szabályozás,¹⁶ azaz az állam szabályozási igényének kiszervezése történt meg.

2. A Facebook moderálási irányelvei avagy a kérdéses tartalmak

A moderálási irányelvek preambulumban a cég a következőképpen fogalmaz: „Komolyan vesszük szerepünket abban, hogy távol tartsuk szolgáltatásunktól a visszaéléseket. [...] Közösségi alapelveinknek mindig is az volt a célja, hogy teret biztosítsanak az embereknek az önkifejezésre és a véleménynyilvánításra.”¹⁷ A CS négy alapértéket azonosít, amit a fentiek során követ, és amelynek betartását minden felhasználójától elvár: hitelesség, biztonság, adatvédelem és méltóság.

¹¹ <https://ncac.org/about-us> (letöltés ideje: 2021. 07. 27.).

¹² NCAC: *Facebook agrees to reconsider artistic nudity policies*, 2019. június 5., <https://ncac.org/news/facebook-agrees-to-reconsider-artistic-nudity-policies> (letöltés ideje: 2021. 07. 27.).

¹³ És valószínűleg félreérti a folyamatot Angelo Stagnaro, amikor azt írja: „A Facebook régóta cenzúrázza a keresztény szervezeteket és egyéneket, meggyőződésünket 'gyűlöletkeltőnek' vagy másként nem megfelelőnek jelöli meg, de a vállalat cselekedetei és a tétlensége az, amit gyűlöletkeltőnek lehetne nevezni.” STAGNARO: i. m.

¹⁴ N/A: *Do naked bodies belong on Facebook? Tech giant struggles with changing 'vague and unevenly enforced' rules over nudity and body art without suppressing freedom of speech*. Daily Mail Online, 2020. január 15., <https://www.dailymail.co.uk/sciencetech/article-7890913/Does-naked-body-belong-Facebook-It-s-complicated.html> (letöltés ideje: 2021. 07. 27.).

¹⁵ Ankush Sinha Roy: *How does Facebook handle the 4+ petabyte of data generated per day? Cambridge Analytica – facebook data scandal*. Medium, 2020. szeptember 16., <https://medium.com/@srank2000/how-facebook-handles-the-4-petabyte-of-data-generated-per-day-ab86877956f4> (letöltés ideje: 2021. 07. 27.).

¹⁶ Arne HINTZ: *Social media censorship, privatized regulation, and new restrictions to protest and dissent = Critical Perspectives on Social Media and Protest: Between Control and Emancipation*, szerk. Lina DENCIK, LEISTERT, Oliver, Rowman & Littlefield, London, 2015, 109–126.

¹⁷ A Facebook magyarul a közösségi alapelvek kifejezést használja, <https://hu-hu.facebook.com/communitystandards/> (letöltés ideje: 2021. 07. 27.).

A CS öt kategóriába sorolja az olyan típusú tartalmakat, amelyeknek nem kíván a saját felületén helyet biztosítani:

1. Erőszak és bűncselekmények
 - a) Erőszak és uszítás
 - b) Veszélyes magánszemélyek és szervezetek
 - c) Összehangolt károkozás és bűncselekmény népszerűsítése
 - d) Szabályozás alá eső termékek
 - e) Csalás és félrevezetés
2. Biztonság
 - a) Öngyilkosság és önsértés
 - b) Gyermekes szexuális kizsákmányolása, gyermekek elleni erőszak és meztelenség
 - c) Felnőttek szexuális kizsákmányolása
 - d) Megfélemlítés és zaklatás
 - e) Emberi kizsákmányolás
 - f) A magánélethez fűződő jogok megsértése és képi megjelenítéshez fűződő személyi jogok
3. Kifogásolható tartalom
 - a) Gyűlöletbeszéd
 - b) Erőszakos és durva tartalom
 - c) Felnőtt meztelenség és szexuális tevékenység
 - d) Szexuális jellegű felhívás
 - e) Kíméletlen és érzéketlen tartalom
4. Integritás és hitelesség
 - a) Kéretlen tartalmak
 - b) Álhírek
 - c) Manipulált média
5. A szellemi tulajdon tisztelete
 - a) Szellemi tulajdon megsértése

Látható, hogy ezek azok a tartalomtípusok, amelyek kapcsán a jogalkotó is sokszor ütközik nehézségekbe. Ki kell ugyanakkor emelni, hogy a tartalomszabályozás „privatizálásának” korszakában e kérdésekről csak és kizárólag a szolgáltatók felelősége volt dönteni, így nem különösképpen meglepő, hogy a CS folyamatosan változik. 2018-ban a vállalat – hosszú idő és sok panasz után – nyilvánossá tette ezeket a moderálási irányelveket: ahogy Pintér Mónika frappánsan megfogalmazta: „*ha a Facebook egy ország, itt az új alkotmánya*”.¹⁸ A nagy változást nem pusztán a nyilvánosság jelentette, hanem az is, hogy megismerhetővé váltak azok a belső irányelvek is, amelyek a cég tartalomszabályozási gyakorlatát irányítják. Ennek alapján a tartalmak osztályozását a mesterséges intelligencia és emberi beavatkozás együttes használatával végzik: ők együtt jelentik az első védvonalat a kéretlen tartalom elleni háborúban.

¹⁸ PINTÉR MÓNIKA: *Ha a Facebook egy ország, itt az új alkotmánya*. 24.hu, 2018. április 24., <https://24.hu/tech/2018/04/24/facebook-uj-szabalyzat-community-standard/> (letöltés ideje: 2021. 06. 27.).

3. Moderálás a mesterséges intelligencia¹⁹ használatával

A mesterséges intelligencia mai formájában még számos nehézséggel nem tud megküzdeni: szövegkontextus, irónia, szlengek stb. „Nagyon jó (az embernél is jobb²⁰) a pornó, a spam és a hamis felhasználói fiókok azonosításában, de még mindig nem jó a gyűlöletbeszéd azonosításában.”²¹ A felhasználók több mint száz nyelven készítenek tartalmakat, és jelen formájában a mesterséges intelligencia még nem képes arra, hogy a bonyolultabb tartalmakat értelmezze. De hogy szükséges, az nem kérdés: nincs az a mennyiségű alkalmazható ember, aki a kétezerhúszas évek adatmennyiségét kezelni tudná. „A Facebook szerint a mesterséges intelligencia eszközei – amelyek közül sokat az emberi moderátorok ’képeznek ki’ – a spam közel 100 százalékát észlelik, és a terroristákkal kapcsolatos eltávolítások 99,5 százalékát, a hamis felhasználói fiókok 98 százalékát, a felnőttek meztelensége és szexuális tevékenysége típusú tartalom 96 százalékát, míg a túlzott erőszakkal kapcsolatos eltávolítások 86 százalékát a mesterséges intelligencia észleli, nem a moderátorok.”²² Mindezen adatokkal szemben a mesterséges intelligencia a gyűlöletbeszédrel kapcsolatos tartalmaknak csak a 36 százalékát volt képes helyesen felismerni, értelmezni és kezelni.²³ Az eljárással kapcsolatban Thiago Dias Oliva arra is felhívja a figyelmet, hogy „amíg a hagyományos bűnüldözés magában foglalja a különböző szereplők által végzett felderítést, büntetőeljárást, elbírálást és büntetést, a mesterséges intelligencia mindezeket egyszerre végzi, kevésbé átlátható módon és kiemelten a korai felismerésre és a megelőzésre összpontosítva.”²⁴ Mindezek mellett Jack M. Balkin – ironikusan – arra a gazdasági racionalitásra is rávilágít, hogy „az algoritmikus alkalmazottak kevesebbe kerülnek, mint az emberi alkalmazottak: nincs családjuk, nem kérnek kávészünetet.”²⁵

Emma Llansó és társai a következőképpen foglalták össze a mesterséges intelligencia használatával kapcsolatos jogalkotási dilemmákat:²⁶

¹⁹ Kiemelendő, hogy bár a közbeszédben a mesterséges intelligencia kifejezést használjuk, ez „utalhat különféle automatizált folyamatok használatára a tartalom moderálásának különböző fázisaiban”. Emma LLANSÓ, JORIS VAN HOBOKEN, PADDY LEERSSEN, JAËN HARAMBAM: *Artificial Intelligence, Content Moderation, and Freedom of Expression*. Working paper of the Transatlantic Working Group on Content Moderation Online and Freedom of Expression, 2020. február 26., <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf> (letöltés ideje: 2021. 06. 27.), 3.

²⁰ Értelmező beszűrés a szerző által.

²¹ Jason KOEBLER, Joseph COX: *The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People*. Vice, 2018. augusztus 23., <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works> (letöltés ideje: 2021. 06. 27.).

²² KOEBLER, COX: *i. m.*

²³ A Facebook és az általa használt mesterséges intelligencia hiányosságairól Joaquin Quiñero Candela, a cég mesterséges intelligenciáért felelős vezetője adott interjút. Karen HAO: *How Facebook got addicted to spreading misinformation*. MIT Technology Review, 2021. március 11., <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> (letöltés ideje: 2021. 06. 27.).

²⁴ Thiago DIAS OLIVA: *Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression*. Human Rights Law Review, 2020/4, 612.

²⁵ Jack M. BALKIN: *Free speech is a triangle*. Columbia Law Review, 2018/7, 2024.

²⁶ LLANSÓ, VAN HOBOKEN, LEERSSEN, HARAMBAM: *i. m.*, 2–3.

– A közbeszédben a mesterséges intelligencia általi moderálást valamilyen más, a technológiát jobban részletező megoldásnak kellene szerepelnie, amely inkább az automatizmusra utal.

– Az automatikus moderálást nem szabad törvényi szabályozás által kötelezővé tenni, mert a technológia jelenlegi állapota se nem kellően megbízható, se nem kellően hatékony.

– Az automatikus moderálást használó tech-cégek szélesebb körű átláthatóságot kell biztosítsanak az eljárásaikhoz.

– Nem lehet mindent automatikus moderálással megoldani.

– Mivel az automatikus moderálás során is alapjogi ütközésekkel szembesülhetünk – leginkább a véleménynyilvánítás szabadsága kérdéskörben –, ezért fontos figyelni, hogy túlzott eltávolítás (over removal) se történjen.

– Nem létezik „semleges” automatikus moderálás.

– A médiaműveltség fejlesztése létfontosságú a kérdés kapcsán.

4. Az ember általi moderálás

Ha az exponenciálisan növekvő tartalom- és felhasználómennyiséget a mesterséges intelligencia nem képes kezelni, akkor az óriás tech-cégek újabb és újabb emberek felvételével válaszoltak. A változás tíz év alatt szinte elképzelhetetlen mértékű: *„A Facebooknál [...] a tartalmak vizsgálata, vitás kérdések eldöntése 2009-ben csupán tizenkét moderátor kezében volt. Ők próbálták igazságos döntést hozni, az akkor százhusz millió felhasználó által közzétett tartalmak és jogösszeütközések esetén.”*²⁷ Vessük ezt össze a 2018-as adattal: *„a Facebook világszerte összesen 7500 moderátort alkalmaz, ebben a pozícióban mostanra 40 százalékkal többen dolgoznak a cégek, mint tavaly ilyenkor.”*²⁸ A Community Operationsnak nevezett munkavállalók azok, akik átnézik a felhasználóktól érkezett jelzéseket a nap huszonnégy órájában, és az esetek döntő többségében huszonnégy órán belül sikerül is feldolgozni a jelentéseket. Munkájukról – a cégek szerint a felhasználók ellen a saját védelmükben – nagyon keveset lehet tudni. A kiszivárgott információk általában szörnyűek (poszttraumás reakciók a megnézett tartalmak miatt, borzasztó munkakörülmények, alacsony bérek, kiszervezett munkavállalók, folyamatos stressz és érzelmi nyomás, mentális problémák, és mindezek miatt elképzelhetetlen mértékű fluktuáció²⁹), amire Mark Zuckerberg annyit mondott: *„Néhány cikk szerintem kissé túl*

²⁷ HUSZÁR Daniella: *A véleménynyilvánítás és kifejezés szabadsága a közösségi média platformjain*. Média Kábel Múhold, 2021/4, 34.

²⁸ N/A: *Nyilvánosak a Facebook moderálási alapelvei*. nlc.hu, 2018. április 24., <https://nlc.hu/szabadido/20180424/facebook-moderalasi-alapelvek-nyilvanos-fellebbezes/> (letöltés ideje: 2021. 06. 27.).

²⁹ Casey NEWTON: *The trauma floor. The secret lives of Facebook moderators in America*. The Verge, 2019. február 25., <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> (letöltés ideje: 2021. 06. 27.).

lett dramatizálva”.³⁰ Ki kell ugyanakkor emelni, hogy a cég „három régióban is egy négy főből álló klinikai pszichológus-csapatot alkalmaz, hogy ezzel segítse a durva, erőszakos tartalommal rendszeresen találkozó moderátorokat”.³¹

Van is, amikor szükség lehet rá: a beszámoló és az adatmennyiség alapján egy-egy bejegyzésről való döntésre nagyjából harminc másodperc jut, és naponta négyszáz körüli esetet lehet megoldani – hogy ennyi idő alatt hogyan lehet értékelni és értelmezni a szöveg kontextusát, minimum kérdéses. Ezzel – Ruckenstein és Turunen szerint³² – az emberi moderátorokat is egyre inkább gépként kezelik a tech-cégek, ami nyilvánvalóan hosszú távon nem lesz kifizetődő a munkakörnyezet szempontjából.

Mivel a moderátorok nem szakértői egy-egy területnek, a Facebook egy igen/nem kérdésekre adható válaszokkal kipreparált sorvezetőt biztosít számukra, amelynek a végén eldönthető, hogy egy tartalom eltávolítandó-e vagy sem. A szövegkontextus, a nyelvek, a nyelvjárások és a felhasználói szándékok tovább bonyolíthatják a kérdést. Ráadásul a jogszabályok jelentősen eltérhetnek országonként, ezért ez a „világsorvezető” nem mindig használható. Itt érünk el Kyle Langvardt javaslatához, aki négy utat vázol fel,³³ amelyen a szabályozó el tud indulni:

a) Az állam expressis verbis meghatározza, hogy milyen tartalmak minősülnek illegálisnak, károsnak vagy eltávolítandónak.

b) A felhasználók választhatnak a szolgáltatók által meghatározott különféle felhasználási módok között, így erősítve az egyéni felelősségvállalást.

c) A tartalom moderálásával kapcsolatos irányelvek és eljárások kötelező közzététele.

d) A tartalommoderálás „privatizálása”, azaz teljes egészében a szolgáltatókra bízva a tartalom közzétételéről való döntést.

Ahogy Kyle Langvardt megjegyzi, egyik út sem jelenti a tökéletes megoldást. A jogi alapkérdés így az: biztosak vagyunk-e abban, hogy a tartalomszabályozásról való döntés privatizálása megfelelő út a jövő számára?

5. Megoldás a moderálás gyakorlatára: együttes használat

A megoldás valószínűleg így nem az emberi moderálás vagy a mesterséges intelligencia általi moderálás közötti döntés, hanem a kettő együttes használatát jelenti majd a jövőben. 2020-ban, a Covid-19-világjárvány alatt – amíg a nagy tech-cégek is otthoni munkavégzésre kötelezték a dolgozóikat, és a mesterséges intelligenciának több feladatot

³⁰ Casey NEWTON: *Read the full transcript of Mark Zuckerberg's leaked internal Facebook meetings*. The Verge, 2019. október 1., <https://www.theverge.com/2019/10/1/20892354/mark-zuckerberg-full-transcript-leaked-facebook-meetings> (letöltés ideje: 2021. június 27.).

³¹ N/A: *Mégis megszólalt a Facebook: így dolgoznak a bejegyzéseinket átnéző moderátorok*. hvg.hu, 2019. július 29., https://hvg.hu/tudomany/20180728_facebook_moderator_bejegyzesek_ellenorzes (letöltés ideje: 2021. 06. 27.).

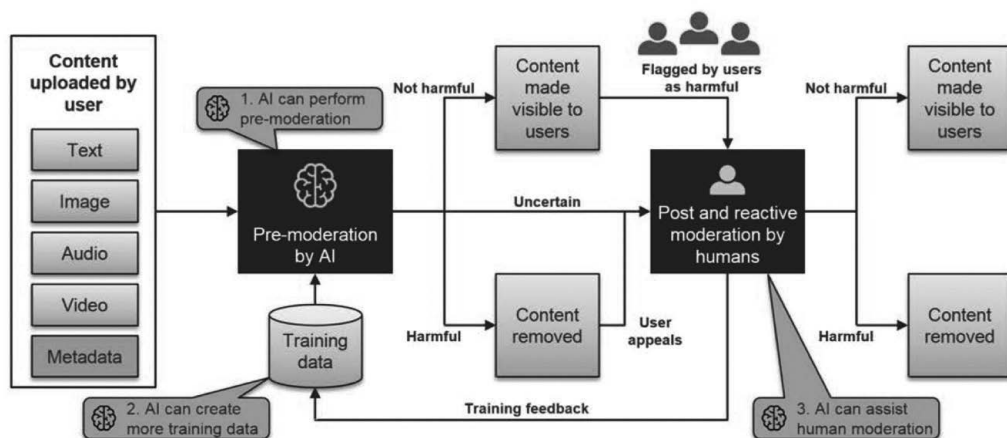
³² Minna RUCKENSTEIN, Linda Lisa Maria RUCKENSTEIN: *Re-Humanizing the Platform: Content Moderators and the Logic of Care*. New Media & Society, 2020/6, 1026–1042.

³³ Kyle LANGVARDT: *Regulating Online Content Moderation*. Georgetown Law Journal, 2017/5, 1353–1388.

adtak³⁴ – „a Facebook és a Google nagyjából megduplázta az eltávolított potenciálisan veszélyes tartalmak mennyiségét az idei második negyedévben a márciusig tartó három hónaphoz képest”,³⁵ és emiatt jóval több panasz is érkezett a döntések ellen, egyértelművé téve, hogy az ember általi tartalomvizsgálat nem nélkülözhető még egy darabig. A saját mesterséges intelligencia-eszközeinek „túlzott cenzúrázási módszereiről” számolt be a YouTube is.³⁶

Az ENSZ-nek a véleménynyilvánítás szabadságáért felelős különmegbízottja jelentést adott ki, amelyben „radikális átláthatóságra”³⁷ szólította fel a közösségi médiaszolgáltatókat a szabályaik megalkotása és betartatása terén. Az együttes használat ráadásul ki tudja szűrni az egyik legveszélyesebb problémát: az elfogultságot. És bár azt gondolhatnánk, hogy ez csak az emberi moderátorokra igaz, az úgynevezett „diszkriminatív algoritmus”³⁸ fogalma is ismeretes, aminek az oka a nem kellő mértékű adatmennyiség, és ami – ebből következően – főleg a közösségi médiában alulreprezentált, kisebbségi csoportokat sújtja.

A moderálással kapcsolatban a lehetséges megoldást a brit médiahatóság, az Ofcom számára készített jelentés így összegzi:³⁹



³⁴ Tarleton GILLESPIE: *Content Moderation, AI, and the Question of Scale*. Big Data & Society, 2020/2, 2.

³⁵ Mark SCOTT, Laura KAYALI: *What happened when humans stopped managing social media content*. Politico, 2020. október 21., <https://www.politico.eu/article/facebook-content-moderation-automation/> (letöltés ideje: 2021. 06. 27.).

³⁶ Alex BARKER, Hannah MURPHY: *YouTube reverts to human moderators in fight against misinformation*. Financial Times, 2020. szeptember 20., <https://www.ft.com/content/e54737c5-8488-4e66-b087-d1ad426ac9fa> (letöltés ideje: 2021. 06. 27.).

³⁷ UNHRC: *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2018, UN Doc A/HRC/38/35, 64.

³⁸ Nicol TURNER LEE, Paul RESNICK, Genie BARTON: *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Brookings Institution, 2019. május 22., <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/> (letöltés ideje: 2021. 06. 27.).

³⁹ CAMBRIDGE CONSULTANTS: *Use of AI in online content moderation*. Ofcom, 2019. július 18., https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf (letöltés ideje: 2021. 06. 27.), 5.

A tanulmány szerint a mesterséges intelligencia használata az alábbi három módon egészítheti ki tökéletesen az ember által végzett moderálást:⁴⁰

– A mesterséges intelligencia felhasználható a moderálás előtti szakasz javítására és a tartalom megjelölésére, emberek általi felülvizsgálatra, növelve a moderálás pontosságát.

– A mesterséges intelligencia segíthet a moderátori felkészülés adatainak összegzésében, ezzel is javítva a moderálás előtti szakasz hatékonyságát.

– A mesterséges intelligencia segíthet az emberi moderátoroknak azáltal, hogy növeli termelékenységüket és csökkenti a tartalom moderálásának potenciálisan káros hatásait az egyes moderátorokra nézve.

Talán mindez hozzájárulhatna ahhoz, hogy a moderálás és a cenzúra kifejezést – valószínűleg újságírói fogásként – ne mossák szándékoltan össze.⁴¹

Irodalomjegyzék

1. Jack M. BALKIN: *Free speech is a triangle*. Columbia Law Review, 2018/7, 2024.
2. Alex BARKER, Hannah MURPHY: *YouTube reverts to human moderators in fight against misinformation*. Financial Times, 2020. szeptember 20., <https://www.ft.com/content/e54737c5-8488-4e66-b087-d1ad426ac9fa> (letöltés ideje: 2021. 06. 27.).
3. CAMBRIDGE CONSULTANTS: *Use of AI in online content moderation*. Ofcom, 2019. július 18.,
4. Sarah CASCOE: *After an Eight-Year Legal Battle, Facebook Ends Its Dispute With a French School Teacher Who Posted Courbet's 'Origin of the World'*. artnet, 2019. augusztus 5., <https://news.artnet.com/art-world/facebook-courbet-lawsuit-ends-1616752> (letöltés ideje: 2021. 07. 27.).
5. Aimee DAWSON: *Facebook censors 30,000 year-old Venus of Willendorf as 'pornographic'*. The Art Newspaper, 2018. február 27., <https://www.theartnewspaper.com/news/facebook-censors-famous-30-000-year-old-nude-statue-as-pornographic> (letöltés ideje: 2021. 07. 27.).
6. Thiago DIAS OLIVA: *Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression*. Human Rights Law Review, 2020/4, 612.
7. Tarleton GILLESPIE: *Content Moderation, AI, and the Question of Scale*. Big Data & Society, 2020/2, 2.
8. Arne HINTZ: *Social media censorship, privatized regulation, and new restrictions to protest and dissent = Critical Perspectives on Social Media and Protest: Between Control and Emancipation*, szerk. Lina DENCIK, Oliver LEISTERT, Rowman & Littlefield, London, 2015, 109–126.
9. Karen HAO: *How Facebook got addicted to spreading misinformation*. MIT Technology Review, 2021. március 11., <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> (letöltés ideje: 2021. 06. 27.).
10. HUSZÁR Daniella: *A véleménynyilvánítás és kifejezés szabadsága a közösségi média platformjain*. Média Kábel Műhold, 2021/4, 34.

⁴⁰ Uo., 6–8.

⁴¹ „[...] annak próbáltunk utánajárni, hogyan is működik a Facebook-cenzúra moderálás [...]” SZILVAY Gergely: *Hogyan és miért cenzuráz a Facebook?* Mandiner, 2018. március 13., https://mandiner.hu/cikk/20180313_hogyan_miért_cenzuraz_a_facebook (letöltés ideje: 2021. 06. 27.).

11. Timothy KARR: *How Censoring Facebook Affects the Fight for Black Lives*. The Root, 2016. augusztus 29., <https://www.theroot.com/how-censoring-facebook-affects-the-fight-for-black-live-1790856542>
12. Jason KOEBLER, Joseph COX: *The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People*. Vice, 2018. augusztus 23., <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works> (letöltés ideje: 2021. 06. 27.).
13. Kyle LANGVARDT: *Regulating Online Content Moderation*. Georgetown Law Journal, 2017/5, 1353–1388.
14. Emma LLANSÓ, Joris VAN HOBOKEN, Paddy LEERSSEN, HARAMBAM, Jaron: *Artificial Intelligence, Content Moderation, and Freedom of Expression*. Working paper of the Transatlantic Working Group on Content Moderation Online and Freedom of Expression, 2020. február 26., <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf> (letöltés ideje: 2021. 06. 27.), 3.
15. Michael MOORE: *Facebook sharing this cat photo could get you BANNED*. Daily Express, 2016. október 5., <https://www.express.co.uk/life-style/science-technology/717978/facebook-bans-user-sharing-cat-photo> (letöltés ideje: 2021. 06. 27.).
16. Casey NEWTON: *Read the full transcript of Mark Zuckerberg's leaked internal Facebook meetings*. The Verge, 2019. október 1., <https://www.theverge.com/2019/10/1/20892354/mark-zuckerberg-full-transcript-leaked-facebook-meetings> N/A: *Mégis megszólalt a Facebook: így dolgoznak a bejegyzéseinket átnéző moderátorok*. hvg.hu, 2019. július 29., https://hvg.hu/tudomany/20180728_facebook_moderator_bejegyzesek_ellenorze (letöltés ideje: 2021. 06. 27.).
17. Casey NEWTON: *The trauma floor. The secret lives of Facebook moderators in America*. The Verge, 2019. február 25., <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> (letöltés ideje: 2021. 06. 27.).
18. PINTÉR Mónika: *Ha a Facebook egy ország, itt az új alkotmánya*. 24.hu, 2018. április 24., <https://24.hu/tech/2018/04/24/facebook-uj-szabalyzat-community-standard/> (letöltés ideje: 2021. 06. 27.).
19. Ankush Sinha ROY: *How does Facebook handle the 4+ petabyte of data generated per day? Cambridge Analytica - facebook data scandal*. Medium, 2020. szeptember 16., <https://medium.com/@srank2000/how-facebook-handles-the-4-petabyte-of-data-generated-per-day-ab86877956f4> (letöltés ideje: 2021. 06. 27.).
20. Minna RUCKENSTEIN, Linda Lisa Maria TURUNEN: *Re-Humanizing the Platform: Content Moderators and the Logic of Care*. New Media & Society, 2020/6, 1026–1042.
21. Mark SCOTT, Laura KAYALI: *What happened when humans stopped managing social media content*. Politico, 2020. október 21., <https://www.politico.eu/article/facebook-content-moderation-automation/> (letöltés ideje: 2021. 06. 27.).
22. Angelo STAGNARO: *Facebook Censors Image of Santa Kneeling Before Our Lord*. National Catholic Register, 2018. december 26., <https://www.ncregister.com/blog/facebook-censors-image-of-santa-kneeling-before-our-lord> (letöltés ideje: 2021. 07. 27.).
23. Victoria STAPLEY-BROWN: *Paris court hears arguments in Facebook censorship case centering on Courbet's Origin of the World*. The Art Newspaper, 2018. február 2., <https://www.theartnewspaper.com/news/paris-court-hears-arguments-in-facebook-censorship-case-centering-on-courbet-s-origin-of-the-world> (letöltés ideje: 2021. 07. 27.).
24. Nicol TURNER LEE, Paul RESNICK, Genie BARTON: *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Brookings Institution, 2019. május 22., <https://www.brookings.edu/wp-content/uploads/2019/05/algorithmic-bias-detection-and-mitigation.pdf>

- www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ (letöltés ideje: 2021. 06. 27.).
25. UNHRC: *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 2018, UN Doc A/HRC/38/35, 64.
 26. N/A: Denmark: *Facebook blocks Little Mermaid over 'bare skin'*. BBC, 2016. január 4., <https://www.bbc.com/news/blogs-news-from-elsewhere-35221329> (letöltés ideje: 2021. 07. 27.).
 27. N/A: *Do naked bodies belong on Facebook? Tech giant struggles with changing 'vague and unevenly enforced' rules over nudity and body art without suppressing freedom of speech*. Daily Mail Online, 2020. január 15., <https://www.dailymail.co.uk/sciencetech/article-7890913/Does-naked-body-belong-Facebook-It-s-complicated.html> (letöltés ideje: 2021. 07. 27.).
 28. N/A: *Facebook Censors Rubens Paintings For Nudity*. frieze, 2018. július 24., <https://www.frieze.com/article/facebook-censors-rubens-paintings-nudity> (letöltés ideje: 2021. 07. 27.).
 29. N/A: *Nyilvánosak a Facebook moderálási alapelvei*. nlc.hu, 2018. április 24., <https://nlc.hu/szabadido/20180424/facebook-moderalasi-alapelvek-nyilvanos-fellebbezes/> (letöltés ideje: 2021. 06. 27.).
 30. NCAC: *Facebook agrees to reconsider artistic nudity policies*, 2019. június 5., <https://ncac.org/news/facebook-agrees-to-reconsider-artistic-nudity-policies> (letöltés ideje: 2021. 07. 27.).
 31. Facebook Community Standards, III.14. Adult nudity and sexual activity, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity/ (letöltés ideje: 2021. 07. 27.).
 32. https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf (letöltés ideje: 2021. 06. 27.), 5.