

SZÖVEG- ÉS ADATBÁNYÁSZAT ÉS GENERATÍV MESTERSÉGES INTELLIGENCIA – AZ INPUTOLDAL SZERZŐI JOGI KIHÍVÁSAI

I. BEVEZETÉS

Közel hatvan évvel ezelőtt, Philip K. Dick sci-fi-író *A tökéletes fegyver* című kisregényének szereplői így beszéltek Orville-ről, a házi robottársalkodóról: „Hatalmas adatbankja van. Mire számítottál, egy új szonettre? Csak azt tudja visszamondani, ami bele lett táplálva. Csak válogatni tud, kitalálni nem”.² Ez az akkor futurisztikusnak hangzó megállapítás mára valósággá vált a beszélő digitális asszisztensek, mint Siri vagy Alexa, online chatbotok, mint a ChatGPT vagy Google Bard, esetleg ezek kombinációja, a Google Gemini³ formájában. A mesterséges intelligencia (MI) természetesen ennél jóval tágabb körben vált életünk részévé. Az MI-től való függés ilyen fokú növekedése pedig számtalan jogi – így például szerzői jogi – kihívást hozott maga után.

A szerzői jogi kutatások sokáig az MI által vagy annak segítségével generált tartalmak védelmére (illetve ahol ennek relevanciája van, úgy azok nyilvántartásba vételére), valamint a létrehozott tartalmak útján elkövethető jogsértésekre fókuszáltak.⁴ 2022 késő nyaratól azonban – a Copilot, majd a szöveg-kép és szöveg-szöveg MI-generátorok piaci alapú és globális elterjedésével – olyan gyors ütemű és széles körű fejlődésen ment keresztül e terület, hogy a kutatási fókusz is eltolódni látszik. 2024-re gyakorlatilag nincs olyan kreatív (hagyományosan szerzői jogilag védett) tartalom, amit ne lehetne MI segítségével előállítani; vagyis a legégetőbb kérdések immáron nem a generált tartalmak védelmével, hanem a gépi tanuláshoz szükséges forrástartalmak felhasználásával kapcsolatosak.

Legalább három körülmény hatott közre a generatív MI sikerében és az ebből eredő jogi kihívások megugrásában. Az első ezek közül az ilyen alkalmazások tömeges (globális és nyilvános/piaci alapú) használata. A második, hogy a generatív MI-technológiákért a platformgazdaság legnagyobb szereplői (például Meta, Alphabet, Baidu, Apple) felelnek

¹ Egyetemi tanár, Szegedi Tudományegyetem; címzetes egyetemi docens (dosentti), University of Turku (Finnország); vezető kutató, Vytautas Magnus University (Litvánia). E-mail: mezei.peter@szte.hu.

² Philip K. Dick: *A tökéletes fegyver*, Agave Könyvek, Vác, p. 59.

³ Cade Metz: Google Releases Gemini, an A.I.-Driven Chatbot and Voice Assistant. The New York Times, 2024. február 8. (<https://www.nytimes.com/2024/02/08/technology/google-gemini-ai-app.html>).

⁴ Vö.: Péter Mezei: From Leonardo to the Next Rembrandt – The Need for AI-Pessimism in the Age of Algorithms. UFITA, 84. évf. 2. sz., 2020, p. 390–429.

(„platformizáció”).⁵ Vagyis a generatív MI immáron nem kísérleti kutatásokkal, hanem tisztán kapitalista alapokon nyugvó fejlesztésekkel azonosítható. A harmadik pedig az, felidézve Dick gondolatát, hogy ezen algoritmusok tanítása és következőképp a végfelhasználói elégedettség előfeltétele a gigantikus adathalmazok feldolgozása („datafikáció”). Ez a három faktor együttállása egyszerre eredményezte a fejlesztések ütemének gyorsulását és a generatív MI iránti verseny fokozódását.

A technológiai fejlesztésekkel párhuzamosan a jogi válaszlelések világszerte körvonalazódnak látszódnak. Az Európai Unióban (EU) sokáig csak normatöredékek mozaikja volt látható, például a CDSM-irányelv⁶ szöveg- és adatbányászatra vagy a tartalommosztó szolgáltatók felelősségére vonatkozó előírásai; esetleg a digitális szolgáltatásokról szóló rendelet⁷ átláthatósági követelményei révén. Legújabbban az EU az ún. MI-rendelettel (AI Act) igyekszik horizontális és holisztikus válaszokat adni az MI térhódítására.⁸

A jelen esszé szempontjából a CDSM-irányelv által újonnan bevezetett, az „információk, például szövegek, hangok, képek vagy egyéb adatok digitális formában történő, automatizált számítógépes elemzés[ére]”⁹ fókuszáló szöveg- és adatbányászati szabályok érdemelnek figyelmet. Ezek egyike a kutatói társadalmat, a másik bárkit (de főleg a vállalkozásokat) érintheti. Épp ez utóbbi, a CDSM-irányelv 4. cikkében foglalt rendelkezések bírhatnak relevanciával a kereskedelmi generatív MI-alkalmazások fejlesztői számára, így a továbbiakban mi is ezekre fókuszálunk csak.

⁵ Az igazsághoz természetesen hozzátartozik, hogy számos olyan szereplője van a generatív AI világnak, amelyek nem sorolhatók közvetlenül a platformgazdaság domináns szereplői alá. Az OpenAI szoros összefonódása jelenleg a Microsofttal relatív újdonság. Lásd: *James Broughel*: OpenAI Is Now Unambiguously Profit-Driven, And That’s A Good Thing. Forbes, 2023. december 9. (<https://www.forbes.com/sites/jamesbroughel/2023/12/09/openai-is-now-unambiguously-profit-driven-and-thats-a-good-thing/>). Hasonlóképp, az Anthropic is a nagy platformoktól független start-upként kezdte, ám 2023–2024-re az Amazon és a Google is komoly befektetőkké váltak. Lásd: *Devin Coldewey*: Amazon doubles down on Anthropic, completing its planned \$4B investment. TechCrunch, 2024. március 27. (<https://techcrunch.com/2024/03/27/amazon-doubles-down-on-anthropic-completing-its-planned-4b-investment/>); *Krystal Hu*: Google agrees to invest up to \$2 billion in OpenAI rival Anthropic. Reuters, 2023. október 28. (<https://www.reuters.com/technology/google-agrees-invest-up-2-blm-openai-rival-anthropic-wsj-2023-10-27/>). Bármikor hasonló jövő várhat a Mistralra vagy a Stable Diffusionre is.

⁶ Az Európai Parlament és a Tanács (EU) 2019/790 irányelve (2019. április 17.) a digitális egységes piacon a szerzői és szomszédos jogokról, valamint a 96/9/EK és a 2001/29/EK irányelv módosításáról (EGT-vonatkozású szöveg), PE/51/2019/REV/1, HL L 130., 2019. 5. 17., p. 92–125.

⁷ 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) (EGT-vonatkozású szöveg), PE/30/2022/REV/1, HL L 277., 2022. 10. 27., p. 1–102.

⁸ 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) EGT-vonatkozású szöveg, PE/24/2024/REV/1, HL L, 2024/1689, 2024. 7. 12.

⁹ CDSM-irányelv, (8) preambulumbekkezdés.

A 4. cikk szerint a tagállamoknak korlátozás vagy kivétel bevezetése révén lehetővé kell tenniük, hogy bárki szabadon többszörözze a jogszerűen hozzáférhető műveket és kapcsolódó jogi teljesítményeket (ideértve az adatbázisok részeinek a kimásolását is), és hogy az ily módon létrehozott másolatokat megőrizze annak érdekében, hogy a digitálisan hozzáférhető szövegek vagy adatok automatizált analitikai módszerekkel történő elemzése útján új ismeretekre tegyen szert, ideértve mintázatok, tendenciák és összefüggések feltárását.¹⁰ Az új szabad felhasználási lehetőség ellensúlyozására az uniós jogalkotó egy „megfelelő módon – az elektronikus úton nyilvánosan hozzáférhetővé tett tartalom esetében géppel olvasható formában –” kifejezett tiltakozási lehetőséget (jogfenntartást) biztosított a szerzői vagy kapcsolódó jogi jogosultak részére. Ahogy a nemzetközi irodalom hívja: kiléptethetik (opt-out) a műveket a szabad felhasználás keretei közül.¹¹ Végül a 4. cikk szerinti szabad felhasználás ugyancsak összhangban kell, hogy álljon a háromlépcsős teszt követelményeivel.¹²

Habár e szabályok első pillantásra tisztának, talán még logikusnak és kiegyensúlyozottnak is tűnhetnek, számtalan dogmatikai és gyakorlati aggály övezi őket, különösen a generatív MI megjelenése óta. Olyan nyitott kérdések sorakoznak előttünk, mint hogy mit tekinthetünk általános vagy kereskedelmi szöveg- és adatbányászat esetén jogszerűen hozzáférhető tartalmak vagy a géppel olvasható formában kifejezett tiltakozás alatt; milyen terjedelmű (általános vagy egyedi tartalmakra vonatkozó) tiltakozásokat kell/lehet tenni; mikor lehet a tiltakozást érvényesen kifejezni (kizárólag a felhasználást megelőzően vagy akár azt követően is); illetve hogy ki tehet érvényesen ilyen nyilatkozatot (csak a jogosult, egy érdekképviseleti szerv, esetleg a fejlesztők saját, internalizált tiltakozási eljárásai is kiválthatják az előírt joghatást).

Az ilyen nyitott kérdések és a szép számmal elérhetővé vált opt-out modell tükrében valószínű, hogy eljött a CDSM-irányelv újrakalibrálásának az ideje. Ennek apropóján nagyon röviden megvizsgáljuk, hogy a generatív MI-alkalmazások tényleg felhasználják-e (szerzői jogi értelemben véve) a forrásokat gépi tanulás céljából; és amennyiben igen, úgy a generatív MI-fejlesztők eredményesen hivatkozhatnak-e a kereskedelmi szöveg- és adatbányászat korlátozására vagy kivételére. Ezek után feltárjuk az uniós jogi előírások bizonytalanságait, különösen a jogfenntartási rendelkezést illetően. Álláspontunk szerint a CDSM-irányelv 4. cikk (3) bekezdése – és részben a kapcsolódó magyar szabályozás – nem ad sem a jogosultaknak, sem a felhasználóknak megbízható válaszokat a generatív MI-alkalmazások szerzői jogi sorsát illetően. Végső soron úgy látjuk, hogy lépéseket kell tenni annak érdekében, hogy az előírások megfelelhessenek a megváltozott iparági helyzetnek annak érdekében,

¹⁰ A kereskedelmi szöveg- és adatbányászat complex fogalmát illetően lásd együtt a CDSM-irányelv 2. cikk (2) bekezdését, 4. cikk (1)–(2) bekezdését és a (18) preambulumbekendést. Lásd továbbá: 1999. évi LXXVI. törvény a szerzői jogról (Szt.), 33/A. § (2) bekezdés 3. pont és 35/A. §.

¹¹ CDSM-irányelv, 4. cikk (3) bekezdés; Szt. 35/A. § (1) bekezdés b) pont.

¹² *Rossana Ducato, Alain Strowel*: Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review*, 2021. 5. sz., p. 329.

hogy továbbra is egy ember- és alkotócentrikus MI-szabályozás létezhesen az Európai Unió területén.

II. BESZÉLHETÜNK-E SZERZŐI JOGI FELHASZNÁLÁSRÓL GENERATÍV MI-ALKALMAZÁSOK ESETÉN?

Habár a generatív MI technológiai dimenzióinak a megértése nem egyszerű feladat, szerencsére ismertek olyan elemzések, amelyek – parafrázálva a CDSM-irányelvet – „jogászok által olvasható formában” kerültek kidolgozásra.¹³ Ezek tükrében – szerzői jogi oldalról közelítve – alapvetően három érdemi kérdésre kereshetünk választ: sor kerül-e engedély- és díjfizetés-köteles felhasználásra a gépi tanulás egésze során; jogvédett tartalom került-e felhasználásra a gépi tanulás és végső soron a tartalomgenerálás során; illetve melyik ország joga alkalmazandó a potenciális felhasználásokra nézve?

Egyik sem könnyű kérdés, és a jelen tanulmány az alkalmazandó jog dimenziójára nem is kíván részletesen kitérni; az elemzés koherenciájának megtartása végett adottnak vesszük a magyar és az uniós jog alkalmazhatóságát.

A másik két kérdést lehetséges, de nem célszerű külön vizsgálni. Mint azt a generatív MI-technológiákkal kapcsolatos első (amerikai, angol és német) perek is igazolják, a felperesekre mindkét részkérdést illetően bizonyítási kötelezettség hárul. Jól látható azonban, mik a következményei annak, ha valamelyik elemet tekintve nem járnak el kellő alaposággal. Az *Andersen*-ügy felperesei ugyanis elmulasztották annak megjelölését, hogy az alperesek mely műveket használták fel a Stable Diffusion gépi tanulási folyamata során. Az eljáró bíróság erre a hiányra világosan felhívta a figyelmet a keresetlevél jelentős részét elutasító végzésében.¹⁴ Ezzel ellentétben a *The New York Times* egy több mint száz oldalas mellékletet csatolt be annak alátámasztására, hogy az OpenAI által fejlesztett ChatGPT tömegesen többszörözte a folyóiratban megjelent cikkeket.¹⁵

¹³ Lásd például *Paulina Jo Pesch, Rainer Böhme: Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern. Gewerblicher Rechtsschutz und Urheberrecht*, 2023. 14. sz., p. 998–1001; *Andrés Guadamuz: A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs. GRUR International*, 2024. 2. sz., p. 113–115.

¹⁴ *Sarah Andersen v. Stability AI Ltd.*, Case 23-CV-00201-WHO, 2023 WL 7132064, p. 7-8. (N.D. Cal. 30 October 2023) („[T]he alleged infringer’s derivative work must still bear some similarity to the original work or contain the protected elements of the original work.”) Hasonlóképp lásd: *Richard Kadrey et al. v. Meta Platforms*, Case 3:23-cv-03417-VC, p. 2. (N.D. Cal. 20 November 2023) („To prevail on a theory that LLaMAs outputs constitute derivative infringement, the plaintiffs would indeed need to allege and ultimately prove that the outputs ‘incorporate in some form a portion of’ the plaintiffs’ books.”)

¹⁵ *The New York Times Company v. Microsoft Corporation et al.*, plaintiff’s complaint, Case 1:23-cv-11195 (27 December 2023), Exhibit J (<https://nytco-assets.nytimes.com/2023/12/Lawsuit-Documents-dkt-1-68-Ex-J.pdf>). A lista tartalmát az OpenAI nyilvános blogbejegyzésben kritizálta; nemcsak technikai hibának titulálva a „memorizálást”, de egyenesen a bizonyítékok manipulálásával vádolva a felperest. Lásd: *Authors: OpenAI and journalism. OpenAI blog*, 2024. január 8. (<https://openai.com/blog/openai-and-journalism#OpenAI>).

Ugyanakkor az, hogy a gépi tanulás, ideértve az adatkészlet összeállítását, a tartalmak finomhangolását és a mintázatok, tendenciák vagy összefüggések feltárását célzó szöveg- és adatbányászatot, szerzői jogilag releváns tárolást vagy másolást jelent, vagy hogy a forrásművek (input) és a generált tartalom (output) között ok-okozati összefüggés áll fenn, messze nem vehető biztosra.¹⁶ Számos technológiai (rész)tevékenység egyáltalán nem eredményez másolást (például egy szó vagy adat ismétlődéseinek a megszámlálása);¹⁷ más esetekben legfeljebb csak a tanulási adatok jellemzőinek az „előrejelzéséről” beszélhetünk.¹⁸ Az irányadó irodalmi álláspontok ezzel együtt is megerősítik, hogy a szerzői jogilag releváns felhasználás(ok) valamilyen formája jelen van a gépi tanulás során.¹⁹

Végso soron ennek a kérdésnek az eldöntése kizárólag a bíróságokon, esetről-esetre lehetséges. Azt bizonyítani, hogy sor került a jogosult művének a felhasználására, komoly, de nem lehetetlen terhet ró a felperesekre. Empirikus bizonyítékok sora igazolja, hogy a különböző – szoftverködsorokra specializált,²⁰ szöveg-kép²¹ és szöveg-szöveg²² – generatív MI-alkalmazások több-kevesebb mértékben „memorizálják” a forrásadatok egyes elemeit. Ahogy ezen tanulmányok egyike megállapította: a memorizálás aránya jelentősen nő a minta duplikálása esetén, illetve a modell kapacitásának és a modell fejlesztésére használt

¹⁶ *Guadamuz*: i. m. (13), p. 113.

¹⁷ Vö.: CDSM-irányelv, (9) preambulumebekezdés („[e]lfordulhat olyan szöveg- és adatbányászat is, amely nem jár többszörözési cselekménnyel”). Lásd továbbá: *Irimi Stamatoudi*: Text and data mining. In: *Irimi Stamatoudi* (szerk.): *New Developments in EU and International Copyright Law*. Wolters Kluwer, Aalphen-an-Rijn, 2016, p. 1261.

¹⁸ *Nicola Lucchi*: ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. *European Journal of Risk Regulation*, 2023 (first view), p. 11.

¹⁹ *Jan Bernd Nordemann, Jonathan Pukas*: Copyright exceptions for AI training data – will there be an international level playing field? *Journal of Intellectual Property Law & Practice*, 2022. 12. sz., p. 973; *Guadamuz*: i. m. (13), p. 115. Jóval szkeptikusabb véleményt képvisel *Pesch, Böhme*: i. m. (13), p. 1001.

²⁰ *Albert Ziegler*: GitHub Copilot research recitation. Github blog, 2021. június 30. (<https://github.blog/2021-06-30-github-copilot-research-recitation/>); *Daphne Ippolito, Florian Tramèr, Milad Nasr, Chiyuan Zhang, Matthew Jagielski, Katherine Lee, Christopher A. Choquette-Choo, Nicholas Carlini*: Preventing Verbatim Memorization in Language Models Gives a False Sense of Privacy. arXiv:2210.17546v3 [cs.LG], 2023, p. 1–26.

²¹ *Gowthami Somepalli, Vasu Singla, Micah Goldblum, Joans Geiping, Tom Goldstein*: Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, p. 6048–6058; *Nicholas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwal, Florian Tramèr, Borja Balle, Daphne Ippolito, Eric Wallace*: Extracting Training Data from Diffusion Models. *Proceedings of the 32nd USENIX Security Symposium*, 2023, p. 5253–5270.

²² *Nicholas Carlini, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramèr, Chiyuan Zhang*: Quantifying Memorization Across Neural Language Models. arXiv:2202.07646 [cs.LG], 2023, p. 1–19; *Kent K. Chang, Mackenzie Cramer, Sandeep Soni, David Bamman*: Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4. arXiv:2305.00118 [cs.CL], 2023, p. 1–16; *Antonia Karamolegkou, Jiaang Li, Li Zhou, Anders Søgaard*: Copyright Violations and Large Language Models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, p. 7403–7412.

tokenek számának növekedésével.²³ A felhasználás tényét ráadásul egyes fejlesztők (például az OpenAI²⁴) közvetett módon maguk is elismerik, amikor a modelleik fair use-teszt alá illeszthetőségét hangsúlyozzák. Mivel e teszt az amerikai jogban ún. *affirmative defense*,²⁵ ezért az csak akkor hívható segítségül kimentésül, ha a jogsértő felhasználás már igazolást nyert. Igaz, tekinthetünk ezekre az érvekre akár „stratégiai kommunikációként” is, tom-pítva ezáltal a cégekre helyeződő nyomást és növelve a beléjük vetett bizalmat, semmint a jogsértés tényét elismerő nyilatkozatoknak.

Végső soron a jelen tanulmány zárásakor okkal feltételezhetjük azt, hogy a memorizálás létező jelenség, és hogy annak szerzői jogi relevanciája is van. Ezt látszik megerősíteni minden eddig hozott végzés az *Andersen v. Stability AI*-, a *Kadrey v. Meta*- és az angol *Getty Images v. Stability AI-ügyben*,²⁶ ahol az eljáró bírók a felperesi keresetlevelek pontjainak többségét megalapozatlannak találták – kivéve a gépi tanulással kapcsolatos pontokat, amelyek további vizsgálata a közeljövőben várható. Ilyen vizsgálatra az uniós és magyar jog tükrében is nagy szükség van.

III. LÉTEZIK-E VALAMILYEN SZABAD FELHASZNÁLÁS A GÉPI TANULÁS ESETÉN?

Nem csupán az előző alfejezet konklúziója miatt, hanem a CDSM-irányelv 4. cikkében foglalt korlátozás vagy kivétel rendszertani logikájából kifolyólag is valószínűsíthető, hogy engedélyköteles felhasználást (többszörözést, kimásolást, átdolgozást) eredményezhet a gépi tanulás. Amennyiben ez igaz (továbbra is fenntartva, hogy egyes magatartások bizonyosan kívül esnek a szerzői jog keretein), logikusan felmerül a kérdés, hogy létezik-e olyan előírás, amely mentén mentesülhetnek a generatív MI-alkalmazások fejlesztői a jogsértések következményei alól. Bár az alkalmazandó jognak e téren is óriási jelentősége lenne (a legtöbb pert az amerikai jog alapján indították, ami a fair use-tesztet tolhatja a vizsgálat középpontjába²⁷), a jelen tanulmány logikájából következően az uniós és magyar normatív környezetet vizsgáljuk meg tüzetesebben.

Jelenleg három olyan lehetőség tűnik legalább részben reálisnak, amelyek uniós szintű harmonizáción alapulnak. Egyrészt már a szoftverirányelv eredeti, 1991-es változata lehetővé tette a jogszerű használó számára a program működésének a megfigyelését, tanulmá-

²³ *Carlini et al.*: i. m. (21), p. 1. A szerzői jogi irodalomból lásd: *Matthew Sag*: Copyright Safety for Generative AI. *Houston Law Review*, 2023, 2. sz., p. 334; *Pesch, Böhme*: i. m. (13), p. 1004.

²⁴ Vö.: OpenAI, i. m. (15).

²⁵ *Mezei Péter*: Mitől fair a fair? Szerzői művek felhasználása a fair use teszt fényében. *Iparjogvédelmi és Szerzői Jogi Szemle*, 3. (113.) évf. 6. sz., 2008. december, p. 26–68.

²⁶ *Getty Images (US) Inc. et al. v. Stability AI Ltd.*, [2023] EWHC 3090 (Ch), para. 108 & 110.

²⁷ Lásd különösen: *Benjamin L. W. Sobel*: Artificial Intelligence’s Fair Use Crisis. *Columbia Journal of Law & the Arts*, 2017. 1. sz., p. 45–97; *Mark A. Lemley, Bryan Casey*: Fair Learning. *Texas Law Review*, 2021. 4. sz., p. 743–785; *Sag*: i. m. (23), p. 303–313.

nyozását és kipróbálását a program elemeinek alapját képező ötletek és elvek meghatározása céljából.²⁸ Mivel ez a nem MI-specifikus kivétel csak a szoftverpéldányok jogszerű használatára vonatkozik, nem magától értetődő, hogy gépi tanulást segítő, masszív adatgyűjtésen alapuló magatartások nem sértik a jogosult érdekeit.²⁹ Egy másik, ugyancsak izgalmas problémakör a szabad forráskódú szoftverek kereskedelmi célú „újrahasznosítása”. E problémakör relevanciáját épp a *GitHub/Copilot*-ügy igazolja.³⁰ Mivel azonban a generatív MI-alkalmazások többnyire nem a szoftver, hanem az Infosoc-irányelv hatálya alá tartozó tartalmakra támaszkodnak a gépi tanulás során, e rendelkezés relevanciája minimálisnak tekinthető.

Nem úgy az Infosoc-irányelv által kötelezően átültetendővé tett időleges többszörözés esete,³¹ amire még a CDSM-irányelv is kifejezetten felhívja a figyelmet.³² Bár eredendően e norma a gyorsítótárolóban való rögzítés és végső soron az internetes böngészés esetére szolgált volna kimentésként, a technológia folyamatos fejlődésének köszönhetően mára ennél több felhasználási módot is felölelhet. Nem lehet ez alól kivétel a gépi tanulás sem, amely, amennyiben megfelel a norma valamennyi konjunktív feltételének, szabad felhasználásnak minősülhet. Nem magától értetődő, hogy valóban sikerrel igazolható lenne valamennyi feltétel jelenléte. Bizonyítási nehézségekbe ütközhet az „önálló gazdasági jelentőség nélküli” jelleg, valamint a szabályozás *sine qua non* feltétele, az adatmegőrzés időlegessége.³³

Végül harmadikként említhető a CDSM-irányelv által ugyancsak kötelezően átültetendővé tett szöveg- és adatbányászat. E norma tartalma azonban részben újdonsága, részben a technológiai fejlődések tükrében sem tekinthető hézagmentesnek. A jogirodalom számos súlyos aggályra hívta fel a figyelmet a terjedelmét, a rendszertani következetlenségeit, a belső ellentmondásait, az alkalmazott kifejezések jelentését és az üzleti titkokra vonatkozó szabályokkal való ütközését illetően.³⁴

²⁸ Jelenleg lásd: Az Európai Parlament és a Tanács 2009/24/EK irányelve (2009. április 23.) a számítógépi programok jogi védelméről (kodifikált változat) (EGT-vonatkozású szöveg), HL L 111., 2009. 5. 5., p. 16–22, 5. cikk (3) bekezdés; Sztj. 59. § (3) bekezdés.

²⁹ *Ducato, Strowel*: i. m. (12), p. 328.

³⁰ Lásd a szerződésszegésre vonatkozó felperesi indítvány elutasítását kérő alperesi indítványt elutasító bírósági végzést: *Doe 1 et al. v. GitHub, Inc., et al.*, 4:22-cv-06823 (N.D. Cal. May 11, 2023) ECF No. 95, p. 21–22.

³¹ Infosoc-irányelv, 5. cikk (1) bekezdés; Sztj. 35. § (6) bekezdés.

³² CDSM-irányelv, (9) és (18) preambulumbekkezdés.

³³ *Thomas Margoni, Martin Kretschmer*: A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. GRUR International, 2022. 8. sz., p. 691–694; *Pesch, Böhme*: i. m. (13), p. 1006; *Guadamuz*: i. m. (13), p. 117–121.

³⁴ Lásd például *Rossana Ducato, Alain Strowel*: Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘Machine Legibility’. IIC – International Review of Intellectual Property and Competition Law, 2019. 6. sz., p. 661–667; *Sean Flynn, Christophe Geiger, João Pedro Quintais, Thomas Margoni, Matthew Sag, Lucie Guibault, Michael Carroll*: Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action. European Intellectual Property Review, 2020. 7. sz., p. 396–398; *Ducato, Strowel*: i. m. (12), p. 326–331; *Margoni, Kretschmer*: i. m. (32), p. 687–689.

Még akkor is, ha a CDSM-irányelvnek ezeket az előírásait kötelező a tagállamoknak átültetni, jelentős eltéréseket mutathatnak a tagállami normák, kockáztatva ezzel az uniós harmonizáció eredményességét. Csak néhány példát hozva: végső soron korlátozás vagy kivétel a szöveg- és adatbányászat? Hogyan ültetik át a tagállamok a jogszerű hozzáférés követelményét?³⁵ Pontosan milyen keretek között őrizhetők meg a másolatok időlegesen?³⁶ Milyen feltételeket támasztanak a tagállami jogalkotók az opt-out vonatkozásában, és mit tekintenek „géppel olvasható formában” kifejezett jogfenntartásnak? A géppel olvasható formájú jogfenntartás mellett elképzelhető emberek által is olvasható jogfenntartó nyilatkozatok alkalmazása is? Hogyan viszonyulnak a gyakorlatban a jogfenntartó nyilatkozatok a hatásos műszaki intézkedésekhez és jogkezelési adatokhoz?³⁷ Habár utóbbiak teljesen más célt szolgálnak, és a jogalapjuk is más, gyakorlati eredményüket tekintve komoly átfedést mutathatnak a jogfenntartó nyilatkozatokkal. E helyütt nincs lehetőség részletesen bemutatni a tagállami átültetési gyakorlatokat, azonban a reCreating Europe projekt rávilágított arra, hogy a nemzeti átültetések jelentős részben defenzív álláspontot tettek magukévé, vagyis nem nagyon tértek el az irányelv szövegétől.³⁸

A kereskedelmi szöveg- és adatbányászat kapcsán a generatív MI megjelenéséig csekély érdeklődést váltottak ki a jogfenntartással kapcsolatos előírások. A 4. cikk (3) bekezdése világosan az MI,³⁹ ám nem a generatív MI kihívásaira reagált. Az elemzők jelentős része „adottnak” vette az opt-out lehetőségét, anélkül azonban, hogy érdemi vizsgálatnak vetették volna alá annak megvalósíthatóságát. Kész talány, hogy pontosan mit értett a jogalkotó a jogfenntartás és annak részletei alatt: a géppel olvashatóság, a jogfenntartás terjedelme, időpontja és gyakorlója alatt. Mindezek nyomán számos olyan gyakorlati variáns jött létre, amely a nyilatkozat telepítésének a helye (forrásoldal kontra gépi tanulás szintje), időpontja (ex ante kontra ex post) és alanya (jogosult, közvetítő vagy MI-fejlesztő) tekintetében érdemi különbséget mutat.

³⁵ Komoly aggályok merülnek fel a kifejezés azonos alkalmazhatóságát illetően a kutatási és kereskedelmi célú szöveg- és adatbányászat keretében. Lásd: *Thomas Margoni: Saving research: Lawful access to unlawful sources under Art. 3 CDSM Directive?* Kluwer Copyright Blog, 2023. december 22. (<https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/>).

³⁶ *Eleonora Rosati: Copyright in the Digital Single Market – Article-by-Article Commentary to the Provisions of Directive 2019/790.* Oxford University Press, Oxford, 2021, p. 89.

³⁷ Infosoc-irányelv, 6–7. cikk; Sztj. 95–96. §. Empirikus elemzések igazolják, hogy egyes elektronikus kereskedelmi szolgáltatások felhasználási feltételei törekszenek géppel történő olvasási lehetőség nélkül is kizárni a szöveg- és adatbányászatot. Lásd: *Ducato, Strowel: i. m. (34)*, p. 668–676.

³⁸ *Caterina Sganga, Magali Contardi, Pelin Turan, Camilla Signoretta, Giorgia Bucaria, Péter Mezei, István Harkai: Copyright Flexibilities: Mapping and Comparative Assessment of EU and National Sources*, 2023, p. 520–523 (<https://ssrn.com/abstract=4325376>).

³⁹ CDSM-irányelv, (8) preambulumbekkezdés.

IV. A SZÖVEG- ÉS ADATBÁNYÁSZATTAL KAPCSOLATOS JOGFENNTARTÁS GYAKORLATI TAPASZTALATAI

Mielőtt a konkrét opt-out modelleket vizsgálánánk meg, érdemesnek tűnik még egyszer szemügyre venni, hogy milyen konkrét elvárásokat fogalmaz meg a CDSM-irányelv és hazánk vonatkozásában az Sztj. Egyrészt a jogfenntartásnak „kifejezésre” kell jutnia, vagyis a jogosultnak a szöveg- és adatbányászati célú többszörözésre és kimásolásra vonatkozó egyértelmű és kifejezett akaratnyilatkozatot kell tennie. Másrészt a jogfenntartás kifejezésének megfelelő módon kell történnie. A CDSM-irányelv (18) preambulumbekzdése e téren tágra nyitja a kapukat, amikor úgy fogalmaz, hogy „[a]z interneten nyilvánosan hozzáférhetővé tett tartalom esetében az említett jogok fenntartása csak géppel olvasható módszer használata esetén tekinthető megfelelőnek, beleértve a metaadatokat és a weboldal vagy szolgáltatás feltételeit is”.⁴⁰ Sőt, az Sztj. vonatkozó rendelkezésének nyelvtani értelmezése még ennél is tágabb kereteket enged, mikor „elektronikus úton nyilvánosan hozzáférhetővé tett” tartalmakról tesz említést, mivel jogvédett tartalmak nemcsak az interneten keresztül, hanem más elektronikai módszerekkel (pl. STP-protokoll) is elérhetővé tehető. Ezenfelül a szabályozásban foglalt géppel olvasható forma nem zárja ki az emberek általi olvashatóságot; az pusztán arra utal, hogy a bányászatra hagyományosan gépi eszközök, ún. botok segítségével kerül sor, vagyis a bányászat automatizált. Más szóval a géppel olvashatóság érdemi feltétele, hogy a bányászatot megvalósító automatizált folyamatok képesek legyenek különbséget tenni a jogfenntartás hatálya alatt álló és nem álló tartalmak között. Végző soron a szabályozás *télosza* nem más, mint hogy megőrizze a jogosult azon jogát, hogy a szöveg- és adatbányászati célú többszörözéseket és kimásolást maguk valósíthassák meg. Egyéb következtetések az uniós normából nem olvashatók ki. Ezzel szemben az Sztj. egy további érdemi feltételt tartalmaz, s mint látni fogjuk rögtön a IV.1. pontban, ennek fontos jelentősége lesz az opt-out gyakorlatát illetően is. Ez a feltétel nem más, mint hogy a jogfenntartó nyilatkozat megtételére kizárólag „előzetesen” kerülhet sor, vagyis azt megelőzően, hogy a jogvédett tartalmat a bányászbötk többszörözhetnék/kimásolhatnák.

1. A JOGFENNTARTÓ NYILATKOZAT MEGTÉTELÉNEK IDŐPONTJA

A CDSM-irányelv nem foglal állást a tekintetben, hogy a jogfenntartó nyilatkozat kifejezésére kizárólag a szöveg- és adatbányászatot megelőzően (*ex ante*) vagy azt követően (*ex post*) is sor kerülhet. Egymásnak ellentmondó érvek és érdekek okán nagyon nehéz megmondani, hogy melyik értelmezés tűnik helyesnek.

⁴⁰ Mivel a generatív MI-technológiák alapvetően elektronikus/az interneten keresztül elérhető forrásokból építkeznek, a (18) preambulumbekzdés másik fordulata („[m]ás esetekben helyénvaló lehet a jogokat más módon, például szerződéses megállapodás vagy egyoldalú nyilatkozat útján fenntartani”) irreleváns vizsgálatunk tárgya szempontjából.

Egyrészt a szerzői jog alapvető történeti sajátossága az előzetes engedélyezés (opt-in) logikája. Nem szabad figyelmen kívül hagyni, hogy a CDSM-irányelv 4. cikk (3) bekezdése nem új jogosultságot vezetett be – épp ellenkezőleg, egy vagyoni jog korlátozásának vagy kivételének a kereteit vágta vissza egy opcionális kontrollmechanizmus útján. Vagyis a korlát/kivétel korlátja/kivétele az opt-out. A nyilatkozat megtételének elmulasztása pedig annyit jelent, hogy a jogosult elfogadja jogvédett tartalmainak a szabad felhasználását. Rádásul mivel a gépi tanulás során megismert adatok „elfelejtése” – ha egyáltalán lehetséges – rendkívül körülményes folyamatnak hangzik, ezért okkal feltételezhető, hogy egy ex post nyilatkozat indokolatlanul nagy (vagy épp lehetetlen/teljesíthetetlen) kötelezettséget róna az MI-fejlesztőre. Mindebből első olvasatra az következik, hogy kizárólag az ex ante jogfenntartás egyeztethető össze a jogalkotói télosszal.

Másrészről azonban a ördög a részletekben rejlik az egyes modelleket illetően. Épp ezért érdemel különös figyelmet a Spawning.ai⁴¹ három eltérő technológiai megoldása. A Spawning.ai-t szerzői jogi alkotók és gépi tanulásban jártas mérnökök 2022 szeptemberében hozták létre annak érdekében, hogy a jogvédett alkotások engedély nélküli, kereskedelmi célú hasznosítását megnehezítsék.⁴² A fejlesztők első technológiai megoldása az *Have I Been Trained* volt,⁴³ amelynek segítségével bárki ellenőrizhette, hogy vizuális alkotása elérhető volt-e a LAION-5B-n keresztül. Ez utóbbi egy több mint ötmilliárd kép publikus elérési útvonalát (linkjét) tartalmazó adathalmaz, amelyen például a Stability AI által fejlesztett Stable Diffusion is tanult. Ha a jogvédett kép linkje elérhető volt az adathalmazban, a jogosult kérhette a tartalom (link) eltávolítását bármely, számára kényelmes időpontban. Ha az ilyen adathalmazok kereskedelmi hasznosítása nem is a legtipikusabb módja a gépi tanulásnak, egy ilyen ex post eltávolítási gyakorlat összeegyeztethetőnek tűnik a jogszabály szövegével és a jogalkotói szándékkal is. Sőt, már csak azért is indokolatlannak tűnik az ex post jogfenntartás automatikus kizárása a szabályozás hatóköréből, mivel a jogosultak igen gyakran olyan felületeken teszik elérhetővé alkotásaikat, amelyek felett nem gyakorolnak technológiai ellenőrzést, pontosabban ahol nem jogosultak előzetesen jogfenntartó nyilatkozatot tenni.

Ugyanakkor erősen kérdéses, hogy az olyan ex post nyilatkozatok, amelyek nem képesek garantálni a jogosult joggyakorláshoz fűződő érdekét, avagy nem tudják hatékonyan biztosítani a kifogásolt tartalom tanulási adatok köréből történő eltávolítását, megfelelnének az uniós előírásoknak. Ebből a szempontból semmi jelentősége annak, hogy a legnagyobb amerikai fejlesztők (például az OpenAI,⁴⁴ a Stability AI⁴⁵ vagy a Bing⁴⁶) saját, ám önkéntes

⁴¹ Lásd: <https://spawning.ai/>.

⁴² Paul Keller, Zuzanna Warso: Defining Best Practices for Opting Out of ML Training – Open Future Policy Brief #5. Open Future Foundation, 2023, p. 8. (<https://openfuture.eu/publication/defining-best-practices-for-opting-out-of-ml-training/>).

⁴³ Lásd: <https://haveibeen trained.com/>.

⁴⁴ Lásd: https://share.hsforms.com/1_OuT5tfFSpic89PqN6r1CQ4sk30.

⁴⁵ Lásd: <https://stability.ai/faq>.

⁴⁶ Lásd: <https://blogs.bing.com/webmaster/september-2023/Announcing-new-options-for-webmasters-to-control-usage-of-their-content-in-Bing-Chat>.

opt-out mechanizmusokat fejlesztettek 2023 őszén,⁴⁷ mivel ezek a megoldások képtelenek biztosítani a már „megtanult tartalmak elfelejtését”.⁴⁸ Arról már nem is beszélve, hogy ezek az internalizált megoldások nem tekinthetők transzparensnek, így nélkülözik a fejlesztők számonkérésének a lehetőségét.

Mindezek fényében úgy látjuk, hogy alapvetően az ex ante jogfenntartó nyilatkozatok illeszthetők az uniós jog keretei közé. Épp ezért figyelemre méltó, hogy a magyar jogalkotó, európai szinten is egyedülálló módon, kizárólag az ilyen, előzetes jogfenntartó nyilatkozatokat tekinti megengedettnek.⁴⁹ Ezzel ugyanakkor az is világos, hogy a generatív MI-fejlesztők internalizált gyakorlatait *expressis verbis* kizárta a magyar jogalkotó a megengedhető modellek köréből.

2. A JOGFENNTARTÓ NYILATKOZAT MEGTÉTELÉNEK A HELYE

Különösen az imént említett iparági fejlemények tükrében vált égetővé annak a kérdésnek a megválaszolása, hogy ki (kinek a kezdeményezésére) és hol (milyen felületen) jogosult a jogfenntartó nyilatkozatot megtenni. Erre nézve két érdemi lehetőség áll rendelkezésre: a jogosult saját döntésére a forrásoldalon helyez el ilyen nyilatkozatot, vagy a gépi tanulási oldalon az MI fejlesztője által kínált formában juttatja kifejezésre akarátát.

2.1. Forrásoldali jogfenntartás

Géppel olvasható formában jogfenntartó nyilatkozatot a forrásoldalon legkönnyebben a robots.txt fájlban, protokoll elhelyezésével a szerveren, HTTP-válaszkód programozásával vagy a html-tartalomba elhelyezett válaszban lehet tenni. A választási lehetőségek változatossága révén a jogosult tulajdonképpen bármely vagy minden adatbányászügynökkel szemben, és bármely vagy minden jogvédett tartalmára nézve képes lehet jogfenntartó nyilatkozatot tenni.

Ilyen géppel olvasható nyilatkozatokat mára számos szereplő fejlesztett. Az ismertebb és előszeretettel telepített nonprofit megoldások közé tartozik a World Wide Web Consortium

⁴⁷ Más cégek modelljei látványosan nem illeszthetők a CDSM-irányelv keretei közé. A Meta „AI Data Deletion Request Process” névre keresztelt eljárása (<https://www.facebook.com/help/contact/510058597920541>) adatvédelmi, és nem szerzői jogi oldalról közelít a kérdéshez. Lásd: Kate Nibbs: Artists Allege Meta’s AI Data Deletion Request Process Is a ‘Fake PR Stunt’. Wired, 2023. október 26. (<https://www.wired.com/story/meta-artificial-intelligence-data-deletion/>). Más fejlesztők, például az Anthropic, úgy tűnik, hogy nem dolgoztak még ki saját opt-out modellt. Vö.: <https://support.anthropic.com/en/articles/7996868-i-want-to-opt-out-of-my-prompts-and-results-being-used-for-training-models>.

⁴⁸ Reece Rogers: Facebook Trains Its AI on Your Data. Opting Out May Be Futile. Wired, 2023. szeptember 7. (<https://www.wired.com/story/facebook-trains-ai-your-data-opt-out/>); Matteo Wong: Artists Are Losing the War Against AI. The Atlantic, 2023. október 2. (<https://www.theatlantic.com/technology/archive/2023/10/openai-dall-e-3-artists-work/675519/>).

⁴⁹ Sztj. 35/A. § (1) bekezdés b) pont.

(W3C) „TDM Reservation Protocol”-ja⁵⁰ vagy a Spawning.ai által hozzáférhetővé tett ai.txt protokoll.⁵¹

Az MI fejlesztésében aktív szerepet vállaló piaci szereplők sem maradnak azonban le a téren. Erre lehet példa a Google ex ante „skalázható ellenőrzési” („scalable control”) mechanizmusa, a Google Extended⁵² vagy az OpenAI információszerzést ex ante kizáró megoldása.⁵³ Végző soron valamennyi itt említett megoldás megfelelni látszik az uniós szabályoknak.

A Chicagói Egyetem két további figyelemre méltó fejlesztéssel állt elő a közelmúltban. Az elsőként elérhetővé tett Glaze a jogosult által kiválasztott vizuális tartalmak „maszkolását” végzi oly módon, hogy az adatbányászügynök a tartalom „fedésére” szolgáló, metaadatokat tartalmazó „stíluskabát” („style cloak”), és nem a kép jellemzőit tanulja meg.⁵⁴ A Glaze folytatásaként a kutatók egy „adatmérgezési” technológiát is kifejlesztettek, melyet Nightshade névre kereszteltek. A kutatócsoport szerint ilyen, a kép tartalmát szándékosan tévesen leíró metaadatok csekély számának a megtanulása is rontja a gépi tanulás határfokát.⁵⁵ Bármennyire is izgalmasak ezek a fejlesztések, nem tűnnek beilleszhetőnek a jogfenntartó nyilatkozatok körébe. Egyrészt nem a forrásmű vagy a nyilatkozat gépi olvasását biztosítják, hanem a maszkolt vagy mérgezett alternatív metaadatok megtanulását segítik elő. Másrészt nem is a jogfenntartó nyilatkozat kifejezésén, hanem az MI-fejlesztők megtévesztésén van a hangsúly. Ezek a megoldások épp ezért – potenciálisan, de nem bizonyosan – csupán hatásvisszaesetnek tekinthetők.

2.2. Gépi tanulási oldali jogfenntartás

A gépi tanulási oldalon kifejezésre juttatott nyilatkozatok összhangja az uniós joggal kételkedés. Egyrészt ezek az egyes MI-fejlesztők egyedi és eseti megoldásai, amelyek arra kényszerítik a jogosultakat, hogy potenciálisan valamennyi fejlesztő tanulási adathalmazát áttekinthetők annak megismerése céljából, hogy az adott algoritmus megtanulta-e a jogvédett alkotásait. Annak tükrében, hogy számos MI-fejlesztő máig nem hozta nyilvánosságra a gépi tanulóhoz használt adathalmazait, a jogosultak imént említett megismerési lehetősége

⁵⁰ Lásd: <https://www.w3.org/2022/tdmrep/>.

⁵¹ Lásd: <https://spawning.ai/ai-txt>.

⁵² Lásd: <https://blog.google/technology/ai/an-update-on-web-publisher-controls/>. Lásd továbbá: Emma Roth: Google adds a switch for publishers to opt out of becoming AI training data. The Verge, 2023. szeptember 28. (<https://www.theverge.com/2023/9/28/23894779/google-ai-extended-training-data-toggle-bard-vertex>); Paul Keller: Generative AI and copyright: Convergence of opt-outs? Kluwer Copyright Blog, 2023. november 23. (<https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/>).

⁵³ Lásd: <https://platform.openai.com/docs/gptbot> és <https://platform.openai.com/docs/plugins/bot>.

⁵⁴ Shawn Shan, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, Ben Y. Zhao: Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models. arXiv:2302.04222v5 [cs.CR], 2023, p. 1.

⁵⁵ Shawn Shan, Wenxin Ding, Josephine Passananti, Haitao Zheng, Ben Y. Zhao: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. arXiv:2310.13828v1 [cs.CR], 2023, p. 1.

is korlátozott vagy épp lehetetlen. Ez pedig a jogok érdemi gyakorlását megnehezíti, ad absurdum megalapozatlan jogfenntartó nyilatkozatok megtételét eredményezheti.

Másrészt számos MI-fejlesztő által elérhetővé tett modell nem fogalmaz világosan a tekintetben, hogy a jogosultnak egyesével kell kezdeményeznie a tartalmak eltávolítását, vagy valamennyi művét kivonhatja a gépi tanulás alól. Ahogy az MI-fejlesztők előszeretettel szoktak érvelni amellet, hogy számukra képtelenség lenne valamennyi tartalom egyedi jogosítása, úgy azt is nehezen látjuk elképzelhetőnek, hogy a fejlesztők egyedileg távolítsanak el tartalmakat a tanulási adathalmazokból. Végző soron pedig, ha a IV.1. alfejezetben foglalt következtetéseink helytállóak, úgy az ilyen kétséges eredményességet mutató, ex post hatályú nyilatkozatok nem felelnek meg a jogalkotói szándéknak.

Ezúttal is hozható azonban kivétel. A Spawning.ai harmadik fejlesztése a Spawning API névre hallgat.⁵⁶ Ennek a segítségével a Spawning.ai közvetlenül az MI-fejlesztőknek teszi lehetővé, hogy az összegyűjtött jogfenntartási nyilatkozatokat naprakészen beépítsék a modelljeikbe (praktikusan kiszűrjék a LAION-5B inkriminált linkjeit). Bár a Spawning.ai saját adatai szerint már több mint másfél milliárd ilyen link eltávolítását kérték a jogosultak,⁵⁷ korlátozott ennek a megoldásnak a gyakorlati haszna, mivel a legnagyobb MI-fejlesztők közül egyelőre csak a Stability AI vállalta önként a Spawning API alkalmazását.⁵⁸

3. ÁLTALÁNOS VAGY EGYEDI JOGFENNTARTÁS?

Harmadikként arra a kérdésre keressünk választ, hogy milyen terjedelműnek kell lennie egy érvényes opt-out nyilatkozatnak. Kizárólag egyedileg, egy-egy konkrét művet megjelölve lehet ilyen nyilatkozatot tenni, esetleg általánosságban, az adott jogosult egész repertoárjára nézve is? Sőt, gyakorlati jelentősége van annak a kérdésnek is, hogy kizárólag a jogosult tehet ilyen nyilatkozatot a saját repertoárja tekintetében, esetleg egy képviselő (egy közös jogkezelő) valamennyi képviselt jogosult teljes repertoárja tekintetében.⁵⁹

Több okból is úgy tűnik, hogy erre a kérdésre rugalmas választ érdemes adni. Egyrészt az alkotómunka egy folyamat, és komoly félreértésekhez vezethet egy általános jogfenntartó nyilatkozat, amelynek alkalmazhatósága a megtételt követően létrehozott és elérhetővé tett tartalmakra nézve nem magától értetődő. Komoly érvek szólhatnak amellet, hogy miképp

⁵⁶ Lásd: <https://spawning.ai/spawning-api>.

⁵⁷ Uo.

⁵⁸ *Benj Edwards*: Stability AI plans to let artists opt out of Stable Diffusion 3 image training. *ArsTechnica*, 2022. december 15. (<https://arstechnica.com/information-technology/2022/12/stability-ai-plans-to-let-artists-opt-out-of-stable-diffusion-3-image-training/>).

⁵⁹ A jelek szerint ugyanis erre is hozható példa Franciaországból. Lásd: *Brad Spitz*: AI data mining: French music collecting society Sacem opts out (with what consequences?) *Kluwer Copyright Blog*, 2024. január 24. (<https://copyrightblog.kluweriplaw.com/2024/01/25/ai-data-mining-french-music-collecting-society-sacem-opts-out-with-what-consequences/>).

a meghatározatlan számú jövőbeli alkotásra adott felhasználási engedély semmis,⁶⁰ úgy meghatározatlan számú jövőbeli alkotásra nézve jogfenntartó nyilatkozatot sem lehet érvényesen tenni. Egyúttal kellő rugalmasságot kell biztosítani a jogosultnak arra, hogy maga döntse el, hogy adott, jövőben megszülető művére nézve is szeretne-e jogfenntartó nyilatkozatot tenni.

Másrészt természetszerű, hogy a jogosultak nem kizárólag saját felületeiken teszik közzé alkotásaikat. Számos ilyen „külső” weboldal tevékenysége felett nem rendelkeznek ellenőrzési joggal, ahol érvényes jogfenntartó nyilatkozatot sem valószínű, hogy tudnak tenni.

Harmadrészt a jogosultak álláspontja bármikor változhat a szöveg- és adatbányászatot illetően általában véve vagy akár csak egy-egy MI-fejlesztő irányába. Gond nélkül elképzelhető, hogy a jogosult csak adott fejlesztőkkel szemben vagy csak adott művekre nézve kíván jogfenntartó nyilatkozatot tenni/fenntartani. A IV.2. pontban bemutatott technológiai megoldások – különösen a robots.txt, a TDM Reservation Protocol vagy a Google Extended – pedig praktikusán programozhatóvá teszi a részleges jogfenntartást is.

Különös tekintettel tehát a CDSM-irányelv ez irányú hallgatására, végső soron arra a következtetésre juthatunk, hogy a jogosultak maximális rugalmasság mellett maguk határozhatják meg jogfenntartó nyilatkozatuk terjedelmét, ideértve annak visszavonási lehetőségét is.

V. ZÁRÓGONDOLATOK

Thierry Breton 2023. március 31-én egy Európai Parlamentben feltett kérdésre adott válaszában még kategorikusan elvetette a jogalkotói beavatkozás igényét az AI-/CDSM-irányelv keresztmetszetében.⁶¹ Ahogy arra a jelen esszé is igyekezett rávilágítani, egy év alatt ez az irány egyre kevésbé tűnik indokoltnak. Az ugyanakkor, hogy magát az irányelvet módosítja-e majd az uniós jogalkotó, egyáltalán nem vehető biztosra. Sokkal valószínűbb, hogy annak kiegészítése soft law eszközökkel, például bizottsági iránymutatással érhető el.⁶²

Ezt a következtetést számos jogpolitikai érv támasztja alá. Általában véve szükséges volna a CDSM-irányelv 4. cikk (3) bekezdését pontosítani, fogalmi elemeit megvilágítani, ideértve a jogfenntartó nyilatkozatok visszaható hatályát (elvárt-e az „elfelejtés” az algoritmusok részéről); indokolt-e (mégis) díjazást előírni a jogosultaknak a gépi tanulás rohamos terjedésére tekintettel;⁶³ érdemes lenne szembesülni azzal, hogy a jogfenntartás gyakorlata rendkívül időigényes és költséges; illetve támpontokra volna szükség a jogfenntartás és a

⁶⁰ Sztj. 44. § (1) bekezdés.

⁶¹ Answer given by Mr Breton on behalf of the European Commission, 2023. március 31. (https://www.europarl.europa.eu/doceo/document/E-9-2023-000479-ASW_EN.html).

⁶² Hasonlóan érvel Keller, *Warso*: i. m. (42); Keller: i. m. (52).

⁶³ Vö.: *Martin Senftleben: Generative AI and Author Remuneration*. IIC – International Review of Intellectual Property and Competition Law, 54. évf. 10. sz., 2023, p. 1535–1560.

hatásos műszaki intézkedések, jogkezelési adatok és az időleges többszörözés átfedéseinek tisztázásához.

Épp ezért tűnik indokoltnak, hogy az Európai Bizottság iránymutatást tegyen közzé ebben a kérdésben, megnyitva a kapukat a sztenderdizálás,⁶⁴ a jó gyakorlatok, a magatartási kódexek vagy bármely más, ennél enyhébb adathasznosítási gyakorlatok felé. A Bizottság egyúttal azt is megvizsgálhatná, hogy milyen negatív következményei lehetnek a tömeges opt-outoknak, különösen az algoritmusok hatékonyságának a csökkenése vagy a generált outputok torzulása formájában.⁶⁵

Az, hogy erre a közeljövőben sor kerüljön, nem valószínű, különös tekintettel a 2024-es európai parlamenti választások és az új Bizottság felállításának időigényes jellegére. Ugyancsak fontos, hogy az EU ténylegesen igyekezett „betömni” egyes lyukakat a már emlegetett mozaikszabályozáson. Az AI Act ugyanis kötelezővé teszi a generatív MI fejlesztői számára egy olyan policy elfogadását, amely azonosítja és ténylegesen tiszteletben tartja a jogfenn tartási nyilatkozatokat; egyúttal előírja, hogy a fejlesztők hozzanak nyilvánosságra egy kellően részletes összefoglalót a gépi tanuláshoz használt adathalmazaik tartalmáról.⁶⁶ Mindkét előírás izgalmas – közjogi – kiegészítése a fennálló szerzői jogi szabályanyagnak; ám egyik előírás sem szól az újonnan bevezetett elképzelések gyakorlati megvalósíthatóságáról. Kiemelt figyelmet kell szentelnünk majd a Mesterséges Intelligenciával Foglalkozó Európai Hivatalnak, amely az említett kellően részletes összefoglalót illetően egységes sablon kidolgozására jogosult. Ezenfelül az, hogy az AI Act elfogadása után annak hatálybalépésére is további éveket kell még várnunk, nem nyugtatja meg a jogosultakat a jelenleg is gőzerővel zajló generatív MI-fejlesztéseket illetően.

⁶⁴ Igaz, erre ilyen iránymutatás nélkül is sor kerülhet az iparági szereplők saját döntései nyomán. Lásd: *Mike Isaac*: Meta Calls for Industry Effort to Label A.I.-Generated Content. The New York Times, 2024. február 6. (<https://www.nytimes.com/2024/02/06/technology/meta-ai-standards-labels.html>).

⁶⁵ *Ducato, Strowel*: i. m. (12), p. 330–331.

⁶⁶ Lásd: AI Act, 53. cikk (1) bekezdés *c*) és *d*) pont: „Az általános célú MI-modellek szolgáltatóinak: ... *c*) a szerzői és kapcsolódó jogokra vonatkozó uniós jognak való megfelelésre irányuló politikát kell bevezetniük, és különösen azért, hogy azonosítsák és betartsák az (EU) 2019/790 irányelv 4. cikkének (3) bekezdése szerint kifejezett jogfenn tartást, többek között a legkorszerűbb technológiák révén; *d*) kellően részletes összefoglalót kell készíteniük – az MI-hivatal által rendelkezésre bocsátott sablonnak megfelelően – és közzétenniük az általános célú MI-modell tanításához használt tartalomról.” Lásd továbbá Vö.: *Teresa Nobre, Leander Nielbock*: An AI Christmas Miracle. Communia Association, 2023. december 18. (<https://communia-association.org/2023/12/18/an-ai-christmas-miracle/>).