DOI 10.26521/profuturo/2023/1/13334

Herke Csongor*

Deepfake: áldás vagy átok? Jogi szabályozási szempontok

Deepfake: a Blessing or a Curse? Legal Regulatory Aspects

Absztrakt

A deepfake olyan videó, hang vagy egyéb tartalom (például kép), amely teljesen vagy részben kitalált, vagy egy meglévő, valós tartalom manipulálásával jött létre. Ahogyan az álhírek (fake news) megkérdőjelezik a valós hírek hitelességét, a mélyhamisítás (deepfake) is megkérdőjelezi a valós tartalmak valódiságát. Ugyanakkor a deepfake-nek a sokszor hangoztatott veszélyei mellett számos előnye is van. A tanulmány a deepfake történeti áttekintését követően ezeket az előnyöket és veszélyeket ismerteti, majd a deepfake észlelésére szolgáló eszközök bemutatását követően kitér a lehetséges jogi válaszlépésekre.

Kulcsszavak: *deepfake története, deepfake előnyei, deepfake veszélyei, bosszúpornó, internetes tartalom hamisítása, büntetőjog*

Abstract

A deepfake is a video, audio or other content (e.g. image) that is completely or partially fabricated or created by manipulating existing, real content. Just as fake news calls into question the authenticity of real news, deepfake also calls into question the authenticity of real content. At the same time, deepfake has many advantages in addition to its often mentioned dangers. Following a historical overview of deepfake, the study describes these benefits and dangers, and then discusses possible legal responses after presenting tools for detecting deepfake.

Keywords: *history of deepfake, benefits of deepfake, dangers of deepfake, revenge porn, falsification of internet content, criminal law*

* Dr. Herke Csongor, tanszékvezető egyetemi tanár, Pécsi Tudományegyetem Állam- és Jogtudományi Kar, Büntető- és Polgári Eljárásjogi Tanszék; e-mail: herke.csongor@ajk.pte.hu; ORCID: 0000-0002-5106-339X.

A magyar labdarúgó válogatott 2022. június 7-én Nemzetek Ligája mérkőzést játszott Cesenában Olaszországgal. A mérkőzés előtt egy nappal a *Nemzeti Sport* internetes oldala rövid cikkében[1] arról számolt be, hogy a *Corriere dello Sport* olasz sportlap szerint az olasz hatóságok komoly erőkkel készülnek a vandál magyar drukkerek megfékezésére.[2] Ahogyan a *Nemzeti Sport* olvasóinak összefoglalták az olasz sportlap cikkét, Cesenában több megbeszélést is tartottak az érkező vendégek fogadása kapcsán, mivel az Europol külön figyelmet szentel a magyar szurkolóknak, akik körében korábban *„neonáci transzparensek kihelyezése, homofób viselkedés, tárgyak pályára dobálása, lelátói zavargások és erőszakos cselekmények is megfigyelhetőek voltak”.* A helyi prefektúra erősítést is kért, hogy fel tudjon készülni a magyar szurkolók fogadására.

Az olasz nyelvű cikkhez egy fotót is mellékeltek, amelyen valóban vandál futballdrukkerek láthatók. Ha azonban jobban megnézzük a képet, az ott látható betonfalon olasz feliratok vannak. Ha pedig valamelyik internetes képforráskereső szoftverrel[3] rákeresünk a képre, könnyen rátalálhatunk a fotó eredetijére.[4] Az eredeti kép egy Lazio-meccs előtt készült 2019. május 15-én (több mint három évvel az olasz–magyar válogatott mérkőzést megelőzően), és az olasz vandálok tevékenységét mutatja be. Röviden tehát: az olasz sportlapban egy olyan képpel illusztrálták, hogy milyen huligánok vannak a magyar szurkolók között, amely egy olasz csapat szurkolóiról készült, akik éppen szétverték Rómát és a megfékezésükre kirendelt rendőröket.

Ez csak egy példája annak, hogy a mai világban milyen egyszerű az embereket félrevezetni. És míg az olasz újságot komoly felelősség terheli (hiszen ők szándékosan tették be a képet, tudván, hogy az nemhogy nem a magyar szurkolókat ábrázolja, hanem éppen az olaszokat), a magyar újságíró már csak az olasz lap oldalát másolta be a cikkébe, ezáltal őt már szándékossággal nem vádolhatjuk (hiszen joggal feltételezhette, hogy az egyik leghíresebb olasz sportlap nem követ el ilyen hamisítást).

Amíg a fenti esetnek nincsen különösebb jelentősége, nem okozott közvetlenül, de talán még közvetve sem kárt, addig a deepfake[5] egyéb formáinak már nemcsak polgári jogi, hanem akár bűntetőjogi következményei is lehetnek. Tanulmányomban először áttekintem a deepfake első megjelenési formáit, majd az előnyeit és hátrányait, végül javaslatot teszek a lehetséges jogi lépésekre, mindezt jogtudományi-elemző, normatív módszertannal. A jogtudományi elemzést nehezíti, hogy a deepfake-kel kapcsolatosan alig találunk magyar szakirodalmat, ezért

[1] https://www.nemzetisport.hu/magyar_valogatott/nl-komoly-rendori-erositessel-keszulnek-a-magyarokra-az-olaszok-2896247 (2023. 07. 07.).

[2] https://www.corrieredellosport.it/news/calcio/italia/2022/06/03-93489520/mobilitazione_a_cesena_in_arrivo_1500_tifosi_ungheresi (2023. 07. 07.).

[3] A Chrome-on belül az Image Search Options segítségével a SauceNAO, IQDB, TinEye, Google Search is segíthet ebben.

[4] https://www.sportmediaset.mediaset.it/foto/calcio/atalanta/coppa-italia-scontri-fuori-dallo-stadio_1025456-2019.shtml (2023. 07. 07.) (ld. a 29. sz. képet).

[5] A deepfake nem túl szabatos magyar fordítása „mélyhamisítás”, azonban (ahogyan pl. a „Photoshop”-ot sem szokás lefordítani „Fotóbolt”-ra) a tanulmányban az angol kifejezést használom.

a külföldi szakirodalmi eredményekre kell támaszkodni. A normatív elemzést pedig az nehezíti, hogy (amint az a tanulmányból is kiderül) nemhogy hazánkban, hanem az USA-ban és az Európai Unióban sincs jelenleg kiforrott szabályozás a deepfake-re. Több szabályozási javaslat is született, de ezek végül nem emelkedtek jogszabályi szintre. Éppen ezért a tanulmány elsősorban a deepfake megjelenési formáival, előnyeivel és veszélyeivel foglalkozik, érintve a lehetséges jogi szabályozási formákat.

1. A deepfake képek, hanganyagok és videók első megjelenési formái

A deepfake olyan videó, hang vagy egyéb tartalom (például kép), amely teljesen vagy részben kitalált, vagy egy meglévő, valós tartalom manipulálásával jött létre.[6] Az Európai Unió mesterséges intelligenciáról szóló jogszabályának módosítására irányuló javaslat szerint a 3. cikk (1) bekezdés 44d. pontja definiálja a deepfake fogalmát: *„manipulált vagy szintetikus hang, kép vagy videótartalom, amely megtévesztő módon az eredetiség vagy hitelesség látszatát kelti, és amely úgy ábrázol személyeket, mintha olyasmit mondanának vagy tennének, amit soha nem mondtak vagy tettek, és amelyet mesterséges intelligencián alapuló technikákkal – beleértve a gépi tanulást és a mélytanulást is – állítanak elő”*.[7]

Ahogy Delfino fogalmaz: a deepfake segítségével annak készítői „hamis valóságot” hoznak létre.[8] Magát a kifejezést egy reddit-felhasználónak[9] tulajdonítják, akinek „deepfakes” volt a felhasználóneve.[10] „Deepfakes” számos híres színésznő arcát cserélte 2017 végén pornófilmekben szereplő színészek arcára,[11] mintha ezek a híres színésznők játszottak volna a pornófilmekben. Ebben nem is az volt a különlegesség, hogy kicserélték az eredeti filmekben szereplők arcát, hanem az, hogy mindezt rendkívül gyorsan, egyszerűen, és bárki által szabadon hozzáférhető, nyílt forráskódú tanulási eszközökkel végezték (például TensorFlow).[12]

A deepfake létrehozásának két klasszikus módja van. Az egyik az úgynevezett Generative Adversial Networks (GAN). Itt két hálózatot (generátor és diszkriminátor) használnak fel. A generátor elkészíti a képet (videót, hangot), a diszkriminátor pedig

[6] Van der Sloot, Bart–Wagensveld, Yvette–Koops, Bert-Jaap: Deepfakes: the legal challenges of a synthetic society. *Computer Law & Security Review, 2022/9, 1. Sorbán ehhez a fogalomhoz hozzáteszi az „algoritmusok általi manipulálásával” kifejezést is.* Sorbán Kinga: A bosszúpornó és deepfake pornográfia büntetőjogi fenyegetettségének szükségességéről. *Belügyi Szemle,* 2020/10, 84.

[7] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_HU.html.

[8] Delfino, Rebecca A.: Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery. *Hastings Law Journal,* 2023/2, 298. (DOI: 10.2139/ssrn.4032094).

[9] A Reddit egy közösségi weboldal, amelyen kizárólag a regisztrált felhasználók megoszthatják híreiket, képeiket és cikkeiket az úgynevezett „subredditeken”. https://www.reddit.com/ (2023. 07. 07.).

[10] Meskys, Edvinas–Liaudanskas, Aidas–Kalpokiene, Julija–Jurcys, Paulius: Regulating Deep Fakes: Legal and Ethical Considerations. *Journal of Intellectual Property Law & Practice,* 2020/15, 26.

[11] https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn (2023. 07. 07.).

[12] Goggin, Benjamin: From porn to 'Game of Thronesű': How deepfakes and realistic-looking fake videos hit it big. *Business Insider,* Jun 23, 2019. https://www.businessinsider.com/deepfakes-explained-the-rise-of-fake-realistic-videos-online-2019-6 (2023. 07. 07.).

jelzi, ha szerinte hamisítvány, és visszaküldi a generátornak. Ez a folyamat mindaddig tart, amíg a diszkriminátor valósnak nem tartja az elkészített tartalmat.[13] Lényegében a generátor egy „hamisító", a diszkriminátor pedig egy „detektív".[14] A másik módszer az Autoencoders (AE). Itt is két hálózat van, a kódoló és a dekódoló. A kódoló a bemeneti képet (videót, hangot) kis vektorként kódolja, a dekódoló pedig megpróbálja ezt az eredeti képbe dekódolni.[15]

Elmondhatjuk, hogy a hamisítás szinte egyidős az audiovizuális eszközök megjelenésével. Louis Daguerre 1838-ban készítette az első dagerrotípiát, az első, emberről készült fényképet. Két évvel később, 1840-ben már meg is jelent az első ismert hamis fénykép, amelyet Hippolyte Bayard készített arról, hogy öngyilkos lett, pedig valójában nem.[16]

Az első híres eset, amikor hamisított fényképet használtak fel politikai célokból, az úgynevezett „Tydings affair" az USA-ban 1950-ben.[17] Millard Tydings marylandi szenátor csalásnak és átverésnek nevezte azt, hogy Joseph McCarthy szenátor idején több száz kommunista dolgozott a külügyminisztériumban. Válaszul McCarthy szövetségesei egy hamis fényképet terjesztettek, amelyen a kommunistaellenes Millard Tydings az USA Kommunista Párt vezetőjével, Earl Browderrel találkozik. A valóság az volt, hogy 1950 júliusa előtt Tydings soha nem találkozott Browderrel, és a fényképet két fényképből állították össze: az egyik, 1938-as fényképen Tydings a rádiót hallgatta, egy 1940-es fotón pedig Browder beszédet mondott.[18] A hamisítás elérte célját, mert Tydings elvesztette a választást John Butlerrel szemben.

1983-ban a brit választások előtt a Crass nevű brit anarcho-punk együttes Margaret Thatcher és Ronald Reagen beszédeinek kivonatait illesztette össze úgy, hogy az egy harcias, politikailag káros telefonbeszélgetésnek tűnjön.[19] Ez még úgynevezett „olcsó" vagy „sekély" hamisítás volt („cheapfekes" vagy „shallow fakes"). Ilyen volt az utóbbi időben a Nancy Pelosi házelnökről készült videó is, amit szándékosan lelassítottak, hogy Pelosi részegnek tűnjön.[20] Ahogyan a számológép különbözik a számítógéptől, úgy különbözik ez a kezdetleges, egyszerű hang deepfake a valódi hanghamisítástól. Hasonló eset volt, amikor 2018 augusztusában BuzzFeed a FakeApp[21] segítségével hamisított videót Obama elnökről[22] (aki a videón így mind-

[13] Davis, Rama–Wiggins, Chris–Donovan, Joan: *Deepfake*. Belfer Center, Cambridge, 2020. 8.

[14] Langa, Jack: Deepfakes, Real Consequences: Crafting Legislation to Combat Threats Posed by Deepfakes. *Boston University Law Review,* 2021/101, 764.

[15] Meskys–Liaudanskas–Kalpokiene–Jurcys: i. m., 4.

[16] https://petapixel.com/2012/11/15/the-first-hoax-photograph-ever-shot/ (2023. 07. 07.).

[17] Feeney, Matthew: *Deepfake Laws Risk Creating More Problems Than They Solve*. Regulatory Transparency Project of the Federalist Society, 2021, 2. https://rtp.fedsoc.org/paper/deepfake-laws-risk-creating-more-problems-than-they-solve/ (2023. 07. 07.).

[18] http://hoaxes.org/photo_database/image/the_tydings_affair (2023. 07. 07.).

[19] https://www.aspi.org.au/report/weaponised-deep-fakes (2023. 07. 07.).

[20] https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/ (2023. 07. 07.).

[21] A FakeApp egy bárki által könnyedén hozzáférhető alkalmazás, amelyet 2018 januárjában hoztak létre és a Google nyílt forráskódú TensorFlow gépi tanulási algoritmusát használja fel. Gerstner, Erik: Face/Off: „DeepFake" Face Swaps and Privacy Laws. *Defence Counsel Journal,* 2020/1, 3.

[22] https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed (2023. 07. 07.).

amellett, hogy a Fekete Párduc gonosztevőjével, Killmongerrel ért egyet a világuralmi tervei kapcsán, Trumpról egyszerűen kijelenti, hogy „dipshit"). Ez a fajta hamisítás ma már nem is túlságosan nehéz, egy egyszerű felhasználó el tudja készíteni: a minta hang alapján a gépelt szöveget a kívánt hangon állítják elő a könnyen hozzáférhető programok (pl. Lyrebird[23]).

Az, hogy létezik hamisítás, arra is jó lehet, hogy akár valós képek valódiságát is meg lehessen kérdőjelezni. Erre jó példa az, amikor 1990-ben a *Newsweek*ben megjelent egy cikk, amelyben a kínaiak azt állították, hogy a pekingi mészárlásról készült digitális képek hamisítottak voltak.[24] Ezt nevezik a nemzetközi szakirodalomban a „hazug osztalékának" („The Liar's Divident"): valaki hamis videót készít a valós videó cáfolatára, valódi videóról állítja, hogy deepfake, vagy éppen a büntetőeljárásban a védelem állítja, hogy a bizonyítékok hamisak, holott tudja, hogy nem azok. Azaz éppen azt használja ki valaki, hogy a hamisítás nem ellenőrizhető, és ebből húz hasznot.

A deepfake a legtöbb helyen a pornográfiával összefüggésben jelenik meg. Egyes szerzők szerint az összes deepfake videó 96%-a pornográf,[25] ami különösen azért magas arány, mert a legnépszerűbb egymillió weboldalnak csak 4%-a pornográf.[26] Márpedig a deepfake videók száma a becslések szerint éves szinten 900%-kal nő.[27] Az elsők között 2016-ban készült Noelle Martin ausztrál aktivistáról egy deepfake pornográf kép, amit számos oldalra kitettek.[28] Ez is jó példa arra, hogy hiába tiltják le ezeket a felvételeket, mert ez a kép is a mai napig felbukkan bizonyos webhelyeken. Itt kell ismételten utalni a „Deepfakes" nevű Reddit-felhasználóval kapcsolatosan említett, Gal Gadot arcával készített 2017-es deepfake pornóvideóra is.

A deepfake lehetőségeire, illetve veszélyeire legjobban az a 2018-as videósorozat mutatott rá, amelyben Nicolas Cage „vállalt szerepet" ismert filmekben (Indiana Jones, James Bond stb.).[29] A választás nem volt véletlen, ugyanis Nicolas Cage volt az egyik főszereplője a nálunk *Ál/Arc* címen bemutatott 1997-es filmnek (eredeti címe: *Face/Off*), amelyben Sean Archer FBI-ügynök (John Travolta) egy orvosi beavatkozással arcot cserél a kómába esett bűnözővel, Castor Troyjal (Nicolas Cage). Abban a filmben tehát orvosi beavatkozás kellett az arccseréhez, míg a deepfake technikához csak egy olcsó (sokszor ingyenes) applikációra van szükség (FaceShifter, FaceSwap, DeepFace Lab, Reface, Snapchat, TikTok[30]).

[23] Langa: i. m., 765.

[24] https://www.newsweek.com/when-photographs-lie-206894 (2023. 07. 07.).

[25] Langa: i. m., 766.; Palmiotto, Francesca: *Detecting Deep Fake Evidence with Artificial Intelligence. A Critical Look from a Criminal Law Perspective*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4384122 (2023. 07. 07.) 3. Más szerzők 95%-ot említenek, pl. Van der Sloot–Wagensveld–Koops: i. m., 6.

[26] Sorbán: i. m., 82.

[27] Letzing , John: How to tell reality from a deepfake? *World Economic Forum,* 2021/4, 1.

[28] https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (2023. 07. 07.).

[29] https://www.irishtimes.com/culture/film/nicolas-cage-is-now-being-spliced-into-every-film-ever-made-1.3376456 (2023. 07. 07.).

[30] https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (2023. 07. 07.).

2019-ben megjelent a ZaoApp az iOS Store-ban. Az applikáció segítségével bárki könnyedén készíthet deepfake videót meglévő filmek és a telefonján készített szelfi felhasználásával. Mivel minden különösebb szakértelem nélkül, nyolc másodperc alatt készít az applikáció egy deepfake videót, nem csoda hát, hogy három nap alatt a legtöbbet letöltött alkalmazássá vált Kínában.[31]

Arra, hogy a deepfake-nek lehet hasznos felhasználása is, David Beckhamnek a malária veszélyeire felhívó videója lehet a jó példa. A videón Beckham kilenc nyelven mondja el a *„Malaria Must Die"* kezdeményezés lényegét, és noha a többi nyelven nem ő beszél, a szájmozgását teljesen az adott nyelven elhangzó szöveghez igazították.[32] Innentől tehát csak egy lépés az, hogy a szinkronizált filmeken az adott színész a saját hangján beszél bármelyik nyelven, aminek persze a rajongók jobban örülnének, mint a munkanélkülivé váló szinkronszínészek.[33] Ugyanezt a technikát alkalmazták 2020-ban Indiában, amikor Manoj Tiwari, az indiai kormányzópárt (Bharatiya Janata Party, BJP) jelöltje a kampányfelvételen hindi dialektusban, harjanvi nyelven beszél, hogy a választói jobban megértsék, miközben az eredeti felvétel angolul készült.[34] A *Star Wars* sorozat példája is ismert: míg a fiatalított Carrie Fisher megjelenése a *Rogue One*-ban még 200 millió dollárba került, a színésznő halála után a modern módszerekkel készített dialógusok költségei *Az utolsó Jedik*ben ennek már csak töredékét tették ki. Szintén a deepfake előnyeire mutatnak rá azok a próbálkozások, amikor ismert, de már elhunyt zenészeknek (Elvis Presley, Frank Sinatra),[35] vagy éppen még élő énekeseknek (Jay-Z)[36] jelentek meg dalai: a mesterséges intelligencia (a továbbiakban: MI) a korábbi dalok stílusa alapján hozott létre új dalokat, amelyeket az ismert zenész hangján szólaltattak meg. 2023 júliusában bukkant fel az interneten a *Titanic* című Oscar-díjas film Celine Dion által énekelt főcímdalának az a változata, ahol a dalt Freddie Mercury „énekli", aki a film megjelenése előtt 6 évvel elhunyt.[37] Ezekkel a deepfake termékekkel kapcsolatos erkölcsi és jogi kérdések megítélése már nem annyira egyértelmű, mint a rajongók hozzáállása.

2. A deepfake előnyei és veszélyei

A deepfake lehetséges előnyeit és hátrányait Citron és Chesney tanulmánya[38] alapján az alábbi táblázatban foglaltam össze:

[31] https://www.bbc.com/news/technology-49570418 (2023. 07. 07.).

[32] Ez az úgynevezett „ajakszinkronizálási" („Lip Syncing") technika. Davis–Wiggins–Donovan: i. m., 6.

[33] https://www.youtube.com/watch?v=QiiSAvKJlHo (2023. 07. 07.).

[34] https://www.technologyreview.com/2020/02/19/868173/an-indian-politician-is-using-deepfakes-to-try-and-win-voters/ (2023. 07. 07.).

[35] https://onezero.medium.com/deepfake-music-is-so-good-it-might-be-illegal-c11f9618d1f9 (2023. 07. 07.).

[36] https://www.theverge.com/2020/4/28/21240488/jay-z-deepfakes-roc-nation-youtube-removed-ai-copyright-impersonation (2023. 07. 07.).

[37] https://www.youtube.com/watch?v=H3Sky9lF6ig (2023. 07. 07.).

[38] Chesney, Bobby–Citron, Danielle: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review,* 2019/107, 1753–1819.

Előnyök	Hátrányok	
1. Oktatás 2. Művészet 3. Autonómia 4. Egyéb	***Egyéneknek vagy szervezeteknek okozott kár***	***Társadalomnak okozott kár***
	1. Kizsákmányolás, kihasználás 2. Szabotázs	1. A demokratikus diszkurzus torzulása 2. A választások manipulálása 3. Az intézményekbe vetett bizalom romlása 4. A társadalmi megosztottság súlyosítása 5. A közbiztonság aláásása 6. A diplomácia aláásása 7. A nemzetbiztonság veszélyeztetése 8. Az újságírás aláásása 9. A hazug osztaléka

Mások csak négy csoportba osztják a deepfake tartalmakat (1. pornografikus, 2. politikai kampányokat vagy 3. a tranzakciós költségek csökkentését szolgáló, 4. kreatív és eredeti deepfake). Mindegyiknek vannak előnyei és hátrányai is, és mindegyikre létezhetnek jogi válaszok[39].

2.1. A deepfake előnyei

A deepfake számos lehetőséget nyújt egy modernebb, a klasszikus oktatási rendszertől alapjaiban eltérő tanítási módszer alkalmazásához. Egy hagyományos középiskolai nyelvórán 15-20 diákkal kell foglalkozni a tanárnak, és nem tud tekintettel lenni az eltérő nyelvi szintekre. A diákok egy része az átlagnál jobban halad(na), más részük lemaradásban van (akár szókincs, akár nyelvtan tekintetében). A jó nyelvtanár valahol középen húzza meg a vonalat, aminek az lesz a következménye, hogy a magasabb nyelvi ismeretekkel rendelkező tanulónak nem elég az új információ, míg a lemaradásban lévő még kevésbé tudja követni az anyagot és még jobban lemarad. Egészen más lenne a helyzet egy MI által vezérelt nyelvórával, ahol nemcsak minden egyes tanuló nyelvi szintjéhez igazítottan lehet haladni az anyaggal, de az MI azt is fel tudja mérni, hogy az adott tanuló szókincsben és/vagy nyelvtanban igényel-e fokozott figyelmet, ezáltal jóval hatékonyabban, személyre szabottan tudja támogatni a hallgatókat. Ehhez tud további adalékkal szolgálni a deepfake, amely a tanulók egyéni érdeklődési körére szabott videóanyagokkal segítheti a tanulók felkészülését. A történelem iránt érdeklődő hallgatóknak történelmi jellegű videókon keresztül tanítható meg az anyagot, a sport iránt fogékony hallgatóknak sporttémákkal.

A történelemoktatás lehetőségei is kibővülnének: míg jelenleg csak a 20. és a 21. század vonatkozásában állnak rendelkezésre videó (és kép) szemléltető anyagok, a deepfake lehetőséget nyújtana arra, hogy akár ókori vagy középkori személyi-

[39] Meskys–Liaudanskas–Kalpokiene–Jurcys: i. m., 5.

ségek is életre keljenek, hiszen a meglévő képek, szobrok, egyéb adatok alapján a deepfake segítségével könnyedén előállíthatók olyan filmek, amelyekben ezek a történelmi személyiségek szerepelnek. Híres példa Kennedy elnök „hanganyaga", amelyben azt a beszédet hallhatjuk, amit Dallasban mondott volna el, ha nem ölték volna meg.[40] És akkor még nem is beszéltünk arról, hogy a sokszor feltett elméleti kérdésekre (például mit mondott volna egymásnak Julius Caesar és Napóleon, ha találkozhattak volna) is kaphatnánk fiktív válaszokat. Az egyébként rendelkezésre álló történelmi felvételeket a deepfake segítségével magyarázatokkal lehetne ellátni,[41] interaktív videók segítségével le lehetne játszani történelmi csatákat más körülmények között (például mi lett volna a szabadságharc kimenetele, ha az oroszok nem segítenek az osztrákoknak?). Végül az oktatás körében kell megemlíteni, hogy a deepfake a közérdekű oktatási kampányok hatékonyságán is sokat segíthetne.[42]

A művészet terén csak utalnék a már korábban említett esetekre, amikor már elhunyt előadókat keltettek életre és hoztak létre nevükben újabb műveket. Biztosan sokan lennének kíváncsiak egy olyan virtuális 3D-s koncertre, ahol maga a közel 200 éve elhunyt Beethoven adja elő az Örömódát vagy a 9. szimfóniát, arról nem is beszélve, hogy mekkora érdeklődés lenne, ha a deepfake befejezné Schubert 8. („Befejezetlen") szimfóniáját, és azt maga a szerző mutatná be (virtuálisan) Bécsben. A deepfake-ből sokat profitálhatna a filmművészet is. Az említett módszerek mellett, amikor már halott színészek „vállalhatnak szerepet" újabb filmekben, vagy éppen az idősebb színész jelenne meg fiatalabb szerepkörben (akár úgy is, hogy a rendelkezésre álló korábbi felvételek alapján egy életrajzi filmben maga a színész játszaná a gyermekkorú, fiatalkorú, középkorú és idősebb szerepeket is), a deepfake kitágítaná a színész lehetőségeit is. Így például nem lenne szükség kaszkadőrre, hiszen az olyan jeleneteknél, ahol eddig kaszkadőr játszotta el a szerepet, a deepfake segítségével megoldható lenne a felvétel az eredeti színésszel. A deepfake előnyei között említik általában a paródiák megjelenését, amelyek néha valóban szórakoztatóak, és ha a jó ízlés határain belül maradnak, akár művészeti értéket is tulajdoníthatunk nekik. Itt említhetjük például a 2019-ben Matteo Renzi volt miniszterelnökről készült híres szatírát[43] vagy ugyanebben az évben a Trump amerikai elnökről készült „felvételeket", amelyekben az elnök fejét szándékosan jóval nagyobbnak, bőrét pedig narancssárgának ábrázolták.[44]

Az Európai Unió mesterséges intelligenciáról szóló jogszabályának módosítására irányuló, korábban már említett javaslata az 52. cikk (3) bekezdés 2. albekezdésével kapcsolatosan kimondja, hogy amennyiben a tartalom nyilvánvalóan kreatív, szatirikus, művészi vagy fikciós filmművészeti alkotás, videójáték vizuális anyaga, vagy

[40] https://www.cereproc.com/en/jfkunsilenced (2023. 07. 07.).

[41] Chesney–Citron: i. m., 1760.

[42] Chesney–Citron: i. m., 1761.

[43] https://observers.france24.com/en/20191008-deepfake-video-former-italian-pm-matteo-renzi-sparks-debate-italy (2023. 07. 07.).
https://www.striscialanotizia.mediaset.it/news/questa-sera-a-striscia-un-fuorionda-incredibile-di-matteo-renzi_9709/ (2023. 07. 07.).

[44] https://www.washingtonpost.com/nation/2019/01/11/seattle-tv-station-aired-doctored-footage-trumps-oval-office-speech-employee-has-been-fired/ (2023. 07. 07.).

hasonló mű vagy program részét képezi, a deepfake-kel kapcsolatosan meghatározott átláthatósági kötelezettségek az ilyen létrehozott vagy manipulált tartalom meglétének megfelelő, egyértelmű és látható módon történő, a mű megjelenítését nem akadályozó közlésére, valamint adott esetben az alkalmazandó szerzői jogok közlésére korlátozódnak.

Néha a szatíra a jogalkalmazó szerint túlmegy a határon. Amikor a Negyedik Köztársaság Párt 2014-es kampányvideóján a két, egymással vitázó miniszterelnök-jelölt szövegét egy-egy majom szájába adták,[45] az Alkotmánybíróság méltóságsértőnek találta[46] ezt a közlést. Egyet kell értenem Mrázzal, miszerint az ilyen, kifejezetten szatirikus ábrázolásnak bele kellene férnie a véleménynyilvánítás szabadságába.[47] Márpedig ha egy átlagember felismeri egy deepfake videóról, hogy az szatíra, paródia, akkor abból jó esetben csak annyi kár keletkezhet, mint egy rádiókabaré adásból.[48]

A deepfake komoly segítséget nyújthat sérült, beteg embereknek egy jobb életminőség eléréséhez. Így a bénulás bizonyos formáiban (például amitrófiás laterálszklerózisban (ALS)[49] vagy Duchenne-féle izomdisztrófiában[50]) szenvedők a saját hangjukon tudnak beszélni.[51] De ugyanez igaz lehet általában minden siketnéma személyre. A deepfake technológia alkalmas lehet arra, hogy ha valaki az élete során veszítette el a hangját, akkor a korábbról rendelkezésre álló hangfelvételek alapján egy mikrofon segítségével a saját hangján megszólaljon. Híres példa erre, amikor egy rádiós műsorvezető egészségi okokból elvesztette a hangját, és ezt követően a CereProc szövegfelolvasó program segítségével tudott „megszólalni".[52] De ha valaki eleve némán született, annak az életminőségén is jelentősen javítana, ha a kiválasztott hangon normális kommunikációt tudna folytatni. A testi fogyatékosságban szenvedőkről olyan videók készülhetnének, amelyeken egészséges ember módjára mozoghatnak (akár sporttevékenységben is részt vehetnek). A virtuális térben pedig (ahogyan azt például a 2023-as Oscar-díjas *Minden, mindenhol, mindenkor* című filmben is láthattuk) maga a játékos jelenhetne meg saját testével, hangjával.[53]

A deepfake egyéb előnyei között szokták felsorolni a bűnüldözéshez nyújtott segítséget (például hamis gyermekpornóval lepleznek le egy pederasztát; a tetthelyet reprodukálják a meglévő adatok alapján). Az Európai Unió mesterséges intelligenciáról szóló jogszabályának módosítására irányuló javaslata az 52. cikk (3) bekezdés 2. albekezdése kifejezetten utal arra, hogy a bűnüldöző hatóságok

[45] https://www.youtube.com/watch?v=0r_CIlvteDY (2023. 07. 07.).

[46] 3122/2014. (IV. 24.) AB határozat (ABH 2014, 610–612.).

[47] Mráz Attila: Deepfake, demokrácia, kampány, szólásszabadság. In: Török Bernát–Ződi Zsolt (szerk.): *A mesterséges intelligencia szabályozási kihívásai*. Ludovika Egyetemi Kiadó, Budapest, 2021, 258.

[48] Langa: i. m., 791.

[49] https://www.huffpost.com/archive/ca/entry/lyrebird-helps-als-ice-bucket-challenge-co-founder-pat-quinn-get-his-voice-back_ca_5cd54688e4b07bc729767cdb (2023. 07. 07.).

[50] https://index.hu/techtud/2023/06/18/mesterseges-intelligencia-d-id-csorba-david-hang/ (2023. 07. 07.).

[51] https://www.cereproc.com/ (2023. 07. 07.).

[52] Langa: i. m., 768.

[53] Ez utóbbi azonban azzal a veszéllyel is járhatna, hogy a felhasználóknak létrejönne egy virtuális alteregójuk, ami bizonyos esetekben nem biztos, hogy hasznos következményekkel járna (így pl. ha egy anya elveszti a gyermekét, de tudja, hogy a virtuális térben „létezik" a fia, komoly pszichológiai következményekkel járna, ha azt hinné, hogy a virtuális térben nap mint nap „érintkezhet" a fiával).

használjanak olyan MI-rendszereket, amelyek célja a deepfake felderítése, valamint a használatukhoz kapcsolódó bűncselekmények megelőzése, kivizsgálása és büntetőeljárás alá vonása.

Szintén hasznos lehet a deepfake, ha a résztvevő védelme érdekében el akarjuk titkolni az adott videóanyagban megszólaló személyazonosságát. Jó példa erre, amikor az HBO 2020-as *Üdvözlet Csecsenföldön* című dokumentumfilmjében homoszexuális csecsenek nyilatkoztak, és a deepfake segítségével megváltoztatták a külsejüket, mivel Csecsenföldön az azonos neműek közötti szexuális irányultság akár halálbüntetéssel is járhat.[54] De ide sorolhatók a kiskereskedelemben rejlő óriási lehetőségek is: a deepfake segítségével megoldható, hogy a vevő tíz perc alatt „felpróbálja" a ruhaüzletben lévő összes, őt érdeklő ruhát, napszemüveget, össze tudja hasonlítani, hogy melyik áll neki jobban, és így gyorsabban, könnyebben, alaposabban tudjon választani (akár az otthonában ülve).

2.2. A deepfake veszélyei

Még a kifejezetten hasznos célra kifejlesztett modern technológiák esetén is fennáll a veszélye annak, hogy káros célokra használják fel. Ahogyan az önvezető járművek is felhasználhatók terrorcselekmény végrehajtására vagy kábítószer-kereskedelemre, esetleg szándékos közúti baleset okozására, úgy a deepfake-nek sem csak hasznos felhasználása lehetséges. Különösen igaz ez azért, mert a történeti kitekintésből is kiderül, hogy már az első „mélyhamisítások" sem kifejezetten társadalomra hasznos célból történtek.

Ahogyan a korábbi táblázatból is jól látható, Citron és Chesney a deepfake veszélyeit két fő csoportra osztja: a) az egyéneknek és szervezeteknek okozott károk, és b) a társadalmi károk csoportjára.

ad a) Az egyéneknek és szervezeteknek okozott károk egyik fajtája a kizsákmányolás (kihasználás). A deepfake segítségével „ellopható" egy komplett személyiség, azaz más nevében végezhet valaki banki tranzakciókat, más nevében felléphet egy politikai rendezvényen. Az első leghíresebb, a *Wall Street Journal* által is ismertetett eset a sokak által csak „brit csalásnak" elnevezett, magyar szállal is bíró átverés volt. Az elkövetők 220 000 euró kárt okoztak, amikor egy brit cég alkalmazottját 2019 márciusában (később kiderült, hogy egy osztrák telefonszámról) „felhívta" Herr Kirsch, a társaság német igazgatója, hogy az alkalmazott utaljon át azonnal egy magyar számlára 220 000 eurót. Mivel az alkalmazott jól ismerte a német igazgató hangját, és már korábban is volt rá példa, hogy telefonon kérte tőle utalások teljesítését, gondolkodás nélkül elutalta az összeget (amit aztán az elkövetők azonnal továbbutaltak Mexikóba, és onnan sok egyéb helyre, ezáltal követhetetlenné téve a pénz útját). A német igazgató azonban valójában nem hívta fel az alkalmazottat, hanem a csalást egy deepfake alkalmazás segítségével hajtották végre. Az elköve-

[54] https://www.nytimes.com/2020/07/01/movies/deepfakes-documentary-welcome-to-chechnya.html (2023. 07. 07.).

tők ezután még kétszer próbálkoztak azzal, hogy nem ment át a pénz és utalják át ismét, de ekkor már az alkalmazott gyanút fogott és nem utalt.[55]

A deepfake segítségével létrehozott videók alkalmasak zsarolásra is. Egy komoly üzletember vagy politikus adott helyzetben inkább fizet a zsarolóknak, holott tudja, hogy a videó csak egy hamisított anyag, hiszen ha nem tudja cáfolni a videó tartalmát, akkor jóval nagyobb kár érheti.

Emberrablások esetén sokszor a legnagyobb gondot az jelenti az emberrablók számára, hogy ha időközben az elrabolt személy meghalt, eltűnt vagy éppen valójában nem is rabolták el, akkor hogyan igazolják a megzsarolt hozzátartozóknak, hogy él és náluk van. A deepfake az ő dolgukat is megkönnyítheti, hiszen nemcsak, hogy könnyedén készíthetnek hamis videófelvételeket az elrabolt személyekről, akár a hangján tudnak válaszolni a telefonban, kérve, hogy minél előbb fizessék ki a váltságdíjat.[56] Ahogyan említettem, a videók 95–96%-a, azaz 20 deepfake videóból 19 szexuális tartalmú. Az ezekkel kapcsolatos kérdések meghaladják jelen tanulmány kereteit, így csak utalnék arra, hogy a deepfake könnyen felhasználható a cyberflashing elkövetéséhez,[57] de akár az úgynevezett „bosszúpornóhoz", vagy meglévő pornófelvételen szerepet vállaló személyek kicseréléséhez is.

A szabotázs is igen komoly károkat okozhat mind az egyének, mind egyes szervezetek számára. Komoly csapások mérhetők az ellenfélre, versenytársra a versenyszférában, ideértve a munkahely, a sport, a piac és a politika területét is. Azt, hogy ez mennyire valós veszély, támasztja alá Laremy Tunsil esete.[58] A híres amerikaifutball-játékos 16 millió dollárt vesztett a 2016-os NFL drafton, mert feltörték a Twitter-fiókját és egy olyan videót raktak ki, amelyen kábítószert fogyaszt. Ugyan Tunsil átesett minden drogteszten, és elmondta, hogy ez a videó évekkel korábban készült róla, a videó káros hatásait nem tudta elkerülni. És bár ez az eset egy valós videó feltöltésével történt, el lehet képzelni, hogy milyen károkat lehet okozni a jövőben a közösségi oldalak feltörésével és deepfake videók feltöltésével. A Microsoftnak egy 2009-ben készített felmérése már azt mutatta, hogy a munkaadók 90%-a ellenőrizte a jelentkezőket a közösségi oldalakon, és az állásra jelentkezők elutasításának 77%-a közösségi oldalakon lévő információkon alapult (akár pusztán azon, hogy a közösségi oldalon a jelentkező „nem megfelelő fényképeket" tárol).[59] A lejáratásra számos módszer állhat rendelkezésre: deepfake videók készíthetők a riválisról részegen való ábrázolásától kezdve a boltból lopás vagy éppen trágár kifejezések használatán át a fent már említett drogfogyasztásig. Közös jellemzőjük az, hogy olyan helyzetben ábrázolják az áldozatot, ami az adott tevékenységgel (sport,

[55] https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 (2023. 07. 07.).

[56] https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (2023. 07. 07.).

[57] Ryan-White, Georgina: Cyberflashing and Deepfake Pornography. *Research and Information Service Briefing Paper*. Northern Ireland Assembly, 2022. http://www.niassembly.gov.uk/globalassets/documents/raise/publications/2017-2022/2022/justice/0122.pdf (2023. 07. 07.).

[58] https://www.si.com/nfl/2016/04/29/laremy-tunsil-draft-controversy-ole-miss-dolphins-twitter-instagram-hacked (2023. 07. 07.).

[59] Chesney–Citron: i. m., 1775.

gazdaság, politika) összeegyeztethetetlen, vagy legalábbis időlegesen komoly károkat okozhat.

ad b) A társadalomnak még szélesebb körben lehet kárt okozni, mint az egyéneknek. Ezek közül Citron és Chesney a demokratikus diskurzus torzulásának veszélyét emeli ki elsőként.[60] Ide sorolható az álhírek terjesztése és az információk torzítása mellett az is, amikor a társadalomban egyébként meglévő hiedelmeket tényeknek tűnő hamisításokkal támasztják alá. Így, ha egy ország nem kívánja aláírni a klímaváltozással kapcsolatos egyezményt, vagy éppen nem kíván eleget tenni az ezzel kapcsolatos intézkedési kötelezettségének, elegendő, ha olyan deepfake videókat és képeket terjeszt, amelyekből az látszik, hogy nemhogy csökkenne, hanem még nő is a sarki jégtakaró.[61] Lényegében tehát akár azt is kétségessé teheti, hogy valós probléma-e az éghajlatváltozás.[62]

A társadalmat veszélyeztető deepfake második csoportjába a választások manipulálása tartozik.[63] Ezt a problémát lényegében minden, a deepfake veszélyével foglalkozó szakirodalom érinti, hiszen nemcsak valós veszélyt jelent a választások manipulálása hamisított fényképek, videók és/vagy hangfelvételek útján, de azok következménye felmérhetetlen a társadalom számára. Amint a fentiekben ismertetett McCarthy–Tyidings-ügy kapcsán is kiderült, a politikusok hamar felismerték azt, hogy egy hamisított fénykép hogyan hathat a választás kimenetelére. Könnyen belátható, hogy mekkora hatást gyakorolhat egy kiváló minőségű deepfake videó, ami egy jó időzítés esetén messzemenően befolyásolhatja a választók akaratát. Csak arra kell figyelni, hogy ne legyen túl messze a választás időpontjától, mert akkor az „áldozat" még időben képes tisztázni magát, de túl közel se legyen, hogy még megfelelően befolyásolhassa a választók (különösen a bizonytalan választók) akaratát (el kell találni az úgynevezett „döntési fojtópontot"[64]). A túl korai időzítés hibájába estek például a 2017-es francia választást befolyásolni kívánók, ugyanis Macron időben tisztázni tudta magát a hamis dokumentumokkal szemben.[65] Egy ilyen deepfake támadás egy kisebb országban is komoly károkat okozhat, de nem nehéz elképzelni, milyen világméretű hatása lehet például egy amerikai választás befolyásolásának. A deepfake nem feltétlenül az egész választás eredményének az alakítására alkalmas, de kiélezett helyzetben, a választás véghajrájában akár döntő hatású is lehet.[66]

Az állami intézmények működésének az alapja és elengedhetetlen feltétele a beléjük vetett bizalom erőssége és tartóssága. Ha ezt bármilyen módon gyengítik, az

[60] Chesney–Citron: i. m., 1777.

[61] https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (2023. 07. 07.).

[62] Verstraete, Mark–Bambauer, Derek E.: Ecosystem of Distrust. *First Amendment Law Review,* 2017/16, 144.

[63] Chesney–Citron: i. m., 1778.

[64] Langa: i. m., 773.

[65] Nossiter, Adam–Sanger, David E.–Perlroth, Nicole: *Hackers Came, But the French Were Prepared*. The New York Times, 2017. május 9. https://www.nytimes.com/2017/05/09/world/europe/hackers-came-but-the-french-were-prepared.html (2023. 07. 07.).

[66] Mráz: i. m., 254.

egész állami intézmény működése veszélybe kerül. Citron és Chesney ilyen lehetséges veszélyként említi, ha készülne egy deepfake videó, amelyen FBI-ügynökök éppen megbeszélik, hogy a hatalmukkal visszaélve miként tudnák üldözni Trump volt USA-elnököt, ezzel is csökkentve esélyét a következő elnökválasztáson.[67] Ez azonban nem csak elméletben létező veszély. A Planned Parenthood egy nonprofit szervezet, amely reproduktív és szexuális egészségügyi ellátást, valamint szexuális oktatást nyújt az Amerikai Egyesült Államokban és világszerte.[68] 2015-ben egy abortuszellenes szervezet, a Center for Medical Progress[69] (CMP) több, állítólag titokban rögzített videót tett közzé. A videók széles körű médiavisszhangot váltottak ki, és negatív színben tüntették fel a Planned Parenthoodot, ami alapján olyan törvényjavaslat előterjesztése merült fel, amely megfosztja a szervezetet a szövetségi családtervezési finanszírozástól. A Planned Parenthood a videókról azt állította, hogy azokat megtévesztő módon szerkesztették, hogy megszégyenítsék a szervezetet. Végül a Planned Parenthood mentesült minden gyanú alól,[70] sőt a CMP ellen indult eljárás. Másik híres eset, amikor egy malajziai minisztert azért tartóztattak le, mert egy (általa deepfake-nek tartott) videó tartalma szerint egy másik férfi bevallotta, hogy vele folytatott szexuális kapcsolatot (ami Malajziában nem megengedett).[71] Még a szakértők sem tudták megállapítani, hogy a videó valós vagy hamisított.[72] Mindkét eset igazolja, hogy milyen nagy kárt képes okozni egy deepfake videó, és milyen nehéz az ellene való védekezés. De a deepfake akár magát a jogállamiságot is alááshatja, ha például készítője a bíróság által az ítélkezés során értékelt bizonyítékok valamelyikéről azt állítja, hogy hamis, vagy éppen ellenkezőleg: a bíróság által hamisként értékelt bizonyítékról állítja azt, hogy valós.

A deepfake általában akkor tudja a legnagyobb hatást gyakorolni, ha egy már meglévő gyanút, bizonytalanságot erősít meg. Hiszen ilyenkor van egy olyan háttér-információ, amit egy deepfake videó megerősít, és egyben a korábbi információ is hitelessé teszi a deepfake videót. Különösen igaz ez a társadalomban egyébként is meglévő megosztottságok vonatkozásában:[73] ezeket a megosztottságokat egy jól elhelyezett deepfake tartalom jelentősen erősíteni tudja. A magyar társadalomban régóta jelen van a cigánybűnözéssel kapcsolatos félelem. A hivatalos statisztikai adatok nem tesznek különbséget a terheltek hovatartozása vonatkozásában, ezért hivatalos adatokkal sem alátámasztani, sem cáfolni nem lehet a cigánybűnözéssel kapcsolatos (sok tekintetben alaptalan) hiedelmeket. Sejthetjük, milyen következményei lennének annak, ha megjelenne és a közösségi médián keresztül hamar széles körben elterjedne egy videó, ahol cigányszervezetek vezetői beszélnek arról, hogy támogatni kell a cigányokat a nem cigányokkal szemben elkövetett bűncselekmé-

[67] Chesney–Citron: i. m., 1779.

[68] https://en.wikipedia.org/wiki/Planned_Parenthood (2023. 07. 07.).

[69] https://en.wikipedia.org/wiki/Center_for_Medical_Progress (2023. 07. 07.).

[70] https://www.nytimes.com/2015/08/28/us/abortion-planned-parenthood-videos.html (2023. 07. 07.).

[71] Harwell, Drew: *Top AI researchers race to detect 'deepfake' videos: 'We are outgunned'*. The Washington Post, 2019. június 12. https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ (2023. 07. 07.).

[72] https://www.wired.co.uk/article/how-to-spot-deepfake-video (2023. 07. 07.).

[73] Chesney–Citron: i. m., 1780.

nyekben, majd bemutat „titkos felvételeket", ahol cigányok az elkövetett bűncselekmények után az említett szervezeti vezetőktől megkapják az elkövetésért járó honoráriumot. Az ilyen jellegű deepfake videók cáfolata is igencsak komoly nehézségekbe ütközik. És mivel ezek a videók egy már egyébként is meglévő társadalmi megosztottságon alapulnak, beláthatatlan és sokszor visszafordíthatatlan következményei lehetnek.

Citron és Chesney a társadalomnak okozott lehetséges károk ötödik csoportjaként a közbiztonság aláásását emeli ki.[74] Komoly veszélyek rejlenek abban, ha deepfake videókon olyan álhíreket terjesztenek el, amelyek arra utalnak, hogy természeti vagy emberi katasztrófa fenyeget. Köztudott, hogy természeti katasztrófák esetén az egyik legnagyobb problémát az jelenti a rendfenntartók számára, hogy miközben emberéleteket és javakat kell menteni, meg kell küzdeni a fosztogatókkal is. Ha elterjed a közösségi médiában, hogy egy aszteroida száguld a Föld felé, amely két napon belül mindent megsemmisít, és a hírt az aszteroidáról készült hamis felvételen felül olyan deepfake videóval támasztják alá, miszerint összeültek a világ legnagyobb hatalmainak vezetői, hogy megtárgyalják, mi a teendő (különösen, ha ezt sikerül egy olyan időpontra igazítani, amikor ezek a vezetők valóban találkoznak egymással), ennek olyan beláthatatlan következményei lehetnek világszerte, amiket akkor is nehéz helyrehozni, ha egyébként két nap múlva bebizonyosodik, hogy az egész csak egy deepfake volt. Ez ugyan csak fikció, de hasonló pánikot okozott 2018-ban az USA-ban, amikor a Hawaii Vészhelyzetkezelési Ügynökség egy alkalmazottja a nyilvánosságot minden alap nélkül egy ballisztikusrakéta-támadásra figyelmeztette, külön jelezve, hogy „Ez nem gyakorlat".[75] Noha időben tisztázódott, hogy a riasztás hamis, az Ügynökség két legfelsőbb polgári tisztviselője bejelentette lemondását, egy középszintű vezetőt felfüggesztettek és az alkalmazottat elbocsátották. Ebből is látszik, hogy noha ez a hamis riasztás nem okozott komoly társadalmi kárt (bár okozhatott volna), az egyénekre kifejtett hatása jelentős volt. Azaz a deepfake társadalomra gyakorolt hatása sokszor együtt jár az egyénekre gyakorolt hatással (és fordítva).

Egy következő lehetséges, a társadalomnak okozott kár a diplomáciai kapcsolatok veszélyeztetése, a diplomácia aláásása, illetve a nemzetbiztonság veszélyeztetése.[76] Ezek megvalósulhatnak akár válsághelyzet előidézésével (2017-ben például orosz hackerek a katari emír szájába adott hamis videókkal és hanganyagokkal indították el a Szaúd-Arábia és Katar közti nemzetközi válságot[77]), de a Kormány tekintélye is könnyen aláásható pár deepfake videóval, és az nyomásgyakorlásra is felhasználható. Maga a diplomácia aláásása is magában hordoz nemzetbiztonsági kockázatokat, de a nemzetbiztonság egyéb módon is veszélyeztethető a deepfake segítségével (például a nemzetközi kapcsolatok megzavarásával, hamis harci műveletek megjelenítésével vagy éppen egy folyamatban lévő háború esetén pol-

[74] Chesney–Citron: i. m., 1781.

[75] https://www.npr.org/sections/thetwo-way/2018/01/30/581853255/hawaii-missile-drill-stated-this-is-not-a-drill-resulting-in-false-alert (2023. 07. 07.).

[76] Chesney–Citron: i. m., 1782–1784.

[77] https://www.theatlantic.com/news/archive/2017/06/qatar-russian-hacker-fake-news/529359/ (2023. 07. 07.).

gári áldozatokkal kapcsolatos hamis állítások közlésével). Ugyanakkor nem szabad megfeledkezni arról sem, hogy a deepfake-nek lehetnek nemzetbiztonsági előnyei is. Langa ezek között emeli ki azt, amikor arcfelismerő szoftver észleli a nemzetbiztonsági fenyegetéseket, amelyek ellen ezáltal időben fel lehet lépni.[78]

Végül[79] meg kell említeni, hogy a deepfake komoly károkat okozhat az újságíróknak is.[80] A bevezetőben említett esetben az olasz újságírónak tudnia kellett, hogy egy hamis képpel illusztrálja a magyar szurkolótábor vandalizmusát, hiszen valószínűsíthetően ő maga szedte le az olasz oldalról a képet. Ugyanakkor a magyar újságíró már csak az olasz (egyébként általában megbízhatónak tekintett) lap írására hivatkozik, az abban megjelent képet mutatja be. Amikor aztán kiderül a csalás, az aláássa a magyar újságíró hitelét is, hiszen joggal gondolhatják azt a magyar szurkolók, hogy milyen az a magyar újságíró, aki az egyébként a külföldi vandalizmusáról egyáltalán nem elhíresült magyar szurkolókat befeketít egy, az olasz szurkolókról készült képpel illusztrálva. Nagyon nehéz dolguk lesz a jövőben az újságíróknak, hogy hitelességüket megőrizzék, és ne kerüljenek egy deepfake-csapdába. Éppen ezért is hozott létre a *The Wall Street Journal* külön bizottságot a hamisított tartalmak azonosítására,[81] de ugyanígy fellépett a deepfake ellen a *Reuters*[82] és a *The Washington Post* is.[83] A Google pedig létrehozta a Jigsaw nevű eszközt,[84] amely szintén segítséget nyújt a manipulált tartalmak észlelésében.

A fentiek alapján egyet kell érteni Mrázzal, aki szerint a deepfake legnagyobb veszélye az, hogy elterjedésével a videó-, kép- és hangfelvételekkel szemben támasztott (és ezáltal közvetve a médiumokba vetett) közbizalom megrendül.[85]

3. A lehetséges jogi szabályozás

A deepfake egyik lehetséges ellenszere lehet az észlelő szoftverek alkalmazása. Számos ilyen észlelő szoftver létezik, a teljesség igénye nélkül ide sorolható:

- a Project Maru, amely a hibás képkockákat elemzi;
- a Project Angora, amely megkeresi neten az adott kép eredeti forrását;
- a DARPA által létrehozott Media Forensics észlelő program;
- a Facebook által alkalmazott „Deep Fake Detection Challenge" keretében 2114 résztvevő által létrehozott mintegy 35 ezer modell;
- a Google MI blogja a deepfake észleléséhez;
- a Graphica, amely felméri, hogy a kép MI-alapú-e és szintetikusan generált-e?;

[78] Langa: i. m., 771.
[79] A „hazug osztalékával" korábban már foglalkoztam.
[80] Chesney–Citron: i. m., 1784.
[81] https://digiday.com/media/the-wall-street-journal-has-21-people-detecting-deepfakes/ (2023. 07. 07.).
[82] Feeney: i. m., 11.
[83] https://www.washingtonpost.com/graphics/2019/politics/fact-checker/manipulated-video-guide/ (2023. 07. 07.).
[84] https://medium.com/jigsaw/disinformation-is-more-than-fake-news-7fdd24ee6bf7 (2023. 07. 07.).
[85] Mráz: i. m., 257.

– a Microsoft által kifejlesztett Video Authenticator, amely észleli a manipulált képeket/videókat;
– és a számos egyéb módszer (automatizált detektálás, frekvenciaelemzés, emberi fül egyedi dinamikáján vagy a száj alakján alapuló módszer, páronkénti tanulás, viselkedés, biometrikus adatok stb.).

Ezekkel az észlelő szoftverekkel az a legnagyobb baj, hogy csak csökkentik, de nem szüntetik meg az ártalmakat (nem védik meg az egyéneket; csak konkrét termékekre jók). Az észlelési technológiák csak a deepfake 65%-át tudják kiszűrni, azt is csak hitelességi százalékkal. Leginkább kontextuális szűrésre alkalmasak („ezt mondaná általában az adott személy?"), technikailag nehéz, fizikailag lehetetlen ilyen mennyiségű digitális tartalmat szűrni.[86] Az itt említett 65% az úgynevezett valódi pozitív arány, amikor a detektor helyesen azonosítja a hamisítványt hamisítványként, és a valódit valódiként. Ehhez képest felmerül a hamis pozitív eredmény (a valódi tartalomra azt mondja a detektor, hogy hamis), valamint a hamis negatív eredmény (a hamis tartalomra mondja a detektor, hogy valódi) veszélye. Míg előbbi a bizonyítékok indokolatlan kizárását eredményezheti, utóbbi könnyen vezethet jogsértő ítélethez.[87] És akkor még nem is említettük Iuvenal római költő híres felvetését: *„Quis custodiet ipsos custodes?"*[88] Azaz előbb-utóbb felmerül annak a kérdése is, hogy maga az észlelő szoftver mennyire megbízható.

Ilyen körülmények között a jogi szabályozás nehézségekbe ütközik. Sloot szerint a jogi szabályozásnak komoly akadályai vannak[89]:

– a technológia túl gyorsan fejlődik, ezáltal a technológiaspecifikus szabályok gyorsan elavulnak;
– nehéz szabályozni, hogy mit lehet tenni és mit nem (se a túl tág, se a túl szűk szabályozás nem jó);
– a határokon átnyúló jelleg nehézségei (ami az egyik országban megengedett, az máshol tilos);
– joghatósági problémák (nem lehet egy másik országra a szabályokat rákényszeríteni);
– összetett hálózat követi el a hamisítást, ezért nehéz feltárni az egyes elkövetők szerepét;
– könnyű megkerülni a joghatóságot (ld. VPN);emellett sok esetben nemcsak a jogsértő, hanem a sértett azonosítása is nehézségekbe ütközik. Az nem kérdés például, hogy sértett az, akinek a képét bemásolják egy meglévő pornófelvételbe. De sértett-e az eredeti pornófelvétel szereplője? Ő ugyanis csak ahhoz adta a hozzájárulását, hogy az eredeti pornófelvételt nyilvánosságra hozzák, de ahhoz nem, hogy egy másik ember arcát másolják rá. Egyáltalán az is kérdéses lehet, hogy ki a jogsértő: aki beszerzi a hamisított videóhoz az alapanyagot, aki elkészíti azt, avagy az, aki feltölti? Ha mindegyik, akkor milyen

[86] Van der Sloot–Wagensveld–Koops: i. m., 3.
[87] Palmiotto: i. m., 7.
[88] Ki őrzi az őrzőket? Más fordításban: Ki figyeli a figyelőket?
[89] Van der Sloot–Wagensveld–Koops: i. m., 7–8.

arányú a felelősség?[90] A peer to peer (P2P) fájlmegosztó webhelyek (Napster, Grokster, Morpheus, Kazaa) egy részét ugyan bezárták, de újabbak és újabbak jönnek létre nap mint nap.[91] Ezek pedig szinte lehetetlenné teszik a bűnfelderítést. Az elkövetők ugyanis többnyire anonim módon töltik fel az anyagot, és proxy szervert vagy anonimizáló webböngészőt (például TOR) használnak, ami (amennyiben az elkövetésnek nincs személyes jellege) szinte lehetetlenné teszi a bűnfelderítést.[92] Ilyenkor merülhet fel a nyelvi azonosítás kérdése (úgynevezett „nyelvi ujjlenyomat"[93]): a nyelvész szakértő akár elkövetői profilt is készíthet (kor, nem, etnikai hovatartozás stb.). Mindenkinek megvan ugyanis a maga stílusa (idiolektus).[94]

A büntetőeljárással kapcsolatosan Palmiotto a jogtalan elítélés, vagy éppen ellenkezőleg: a jogtalan felmentés számos veszélyére hívja fel a figyelmet a deepfake-kel összefüggésben.[95] Így a deepfake megjelenése előtt egy, a tetthelyről az elkövetéskor készült videó döntő bizonyíték lehetett a terhelttel szemben, vagy éppen a terhelt számára szolgáltathatott megdönthetetlen alibit. A deepfake korában ezeknek a bizonyítékoknak a bizonyító erejük jelentősen csökkenhet. Nehéz helyzetbe kerülhet a bíró, ha azt látja, hogy a tetthelyen készült felvételen egy, a vádlottra igencsak hasonlító személy látható, ugyanakkor a vádlott bemutat egy videófelvételt, miszerint a bűncselekmény elkövetésének időpontjában egy labdarúgó-mérkőzésen vett részt, és a videón még látszik is a háttérben az eredményjelző, amin jól kivehető az időpont. Ha fel is merül a bíróban a deepfake gyanúja, amíg ezt nem tudja bizonyítani, két videófelvétel áll egymással szemben, és a vádlott joggal állíthatja, hogy az övé az eredeti (márpedig az in dubio pro reo elve alapján azt is kell elfogadni). És elég csak az O. J. Simpson ügyre gondolni, ahol a nyomozók a terhelttől levett vérrel locsolták tele a bűncselekmény helyszínét annak érdekében, hogy bizonyítékot kreáljanak O. J. Simpsonnal szemben.[96] Innen már csak egy lépés az, hogy deepfake videóval, képpel vagy hanganyaggal bizonyítsák egy ártatlan bűnösségét. És nem is feltétlenül a bűnüldöző szerv részéről kell a hamisításnak történnie: egy féltékeny (volt) férj egy jól elkészített deepfake videóval könnyen börtönbe juttathatja (volt) feleségét vagy annak párját stb. És akkor még nem is beszéltünk a „hazug osztalékáról" (amit korábban már említettem): a megvádolt személy akkor is hivatkozhat arra, hogy a bizonyítékok között szereplő videó deepfake, ha maga is tudja, hogy valós. Amíg nem lehet kétséget kizáróan bizonyítani az adott videó valódiságát, egy ilyen hivatkozás könnyen vezethet felmentéshez.

[90] Meskys–Liaudanskas–Kalpokiene–Jurcys: i. m., 6.
[91] Gerstner: i. m., 10.
[92] Sorbán: i. m., 98.
[93] Helyesebben ujjnyom, az ujjlenyomat ugyanis szakmailag helytelen, csak az újságírózsargonban elterjedt kifejezés.
[94] Ürmösné Simon Gabriella–Nyitrai Endre: The phenomena of epidemic crime, deepfakes, fake news, and the role of forensic linguistics. *Információs Társadalom,* 2021/4. 86., 96.
[95] Palmiotto: i. m., 5.
[96] Pap András László: Nyomrögzítés rivaldafényben: O. J. Simpson és az évszázad pere. *Rendészeti Szemle,* 2010/7-8. 226–251.

Már a jelenlegi jogi környezetben is számos lehetőség van a deepfake okozta sérelmek jogi úton történő orvoslására. Ha lehet bizonyítani, hogy az adott videó (hanganyag, kép) deepfake, annak létrehozójával szemben mind a szerzői jogsértés, mind az okozott kár megtérítése iránt polgári jogi per indítható. Ugyanígy perelhető a tartalomszolgáltató is, ha ismételten teszi közzé a jogellenes tartalmat azt követően, hogy felhívták rá a figyelmet, hogy ne tegye, vagy ha kifejezetten a tartalomszolgáltató kéri a deepfake tartalmat, avagy éppen szándékosan módosítja a felületét, hogy ne lehessen észlelni a deepfake tartalmakat (illetve a tárhelyen lévő tartalmak ilyen voltát). A Btk. is számos helyen tartalmaz olyan bűncselekményi tényállásokat, amelyeket a deepfake tartalom létrehozója (közzétevője) megvalósíthat.[97] Ugyanakkor lehetséges olyan tényállások megalkotása is, amelyek kifejezetten a deepfake-hez kapcsolódnak, illetve a meglévő büntetőjogi tényállások minősített esetei bővítése is.

Kifejezetten a deepfake-kel kapcsolatosan merül fel elsőként az a kérdés, hogy a saját célra készített deepfake szexvideókkal kapcsolatosan szükség van-e a kriminalizálásra.[98] Amíg ez a „saját célra" való készítés valóban saját célt szolgál (azaz arról senki nem szerez tudomást), addig a büntetőjogi üldözés szükségtelen (ugyanúgy, mint ahogyan nem üldözzük azt, ha valaki az otthonában ülve a tv képernyője előtt rágalmazó/becsületsértő kifejezést használ mással szemben). Amint azonban ez a videó mások tudomására jut, már felmerül a büntetőjogi felelősség kérdése is. Ehhez pedig nem szükséges, hogy a deepfake videó készítője szándékosan juttassa el a videót másoknak, sőt, álláspontom szerint még gondatlanság sem szükséges ehhez. Azaz, ha valaki deepfake szexvideót készít például egy híres sportoló fényképének felhasználásával, majd feltörik a számítógépét, akkor is felel (polgári és büntetőjogi szempontból egyaránt) a videó elkészítéséért, ha ő maga a videót – állítása szerint – nem kívánta nyilvánosságra hozni. Ugyanakkor a deepfake videók terjedésének megakadályozásához fűződő érdek miatt indokolt lehet büntethetőséget megszüntető vagy korlátlanul enyhíthető okként szabályozni, ha a készítő a videó deepfake jellegét (mielőtt abból komoly következmények származtak volna) feltárja. Le kell szögezni: önmagában a pornográf felvételek készítése és terjesztése nem tiltott (egy egész legális iparág épül erre), sőt (a szólásszabadsággal összefüggésben) kifejezetten alapjogi védelemben is részesül. A pornográfia extrém fajtái viszont kifejezetten üldözendők (például különösen kegyetlen, állatokkal való fajtalankodást bemutató vagy gyermekpornográfia).[99] A deepfake pornográf felvétel (és a bosszúpornó) is ilyen terület, hiszen az áldozat nem járul hozzá az adott felvételen való részvételhez, illetve annak nyilvánosságra hozatalához, ezáltal ez már ütközik

[97] Ilyen pl. a kapcsolati erőszak Btk. 212/A. §, a személyes adattal visszaélés Btk. 219. §, a zaklatás Btk. 222. §, a rágalmazás Btk. 226. §, becsület csorbítására alkalmas hamis hang- vagy képfelvétel nyilvánosságra hozatala Btk. 226/A. §, becsületsértés Btk. 227. §, rémhírterjesztés Btk. 337. §, közveszéllyel fenyegetés Btk. 338. §, a választás, népszavazás és európai polgári kezdeményezés rendje elleni bűncselekmény Btk. 350. §, a zsarolás Btk. 367. §, egyes szellemi tulajdonjog elleni bűncselekmények (bitorlás Btk. 384. §, szerzői vagy szerzői joghoz kapcsolódó jogok megsértése Btk. 385. §, védelmet biztosító műszaki intézkedés kijátszása Btk. 386. §, információs rendszer védelmét biztosító technikai intézkedés kijátszása Btk. 423. §).

[98] Van der Sloot–Wagensveld–Koops: i. m., 8.

[99] Sorbán: i. m., 85–86., 91.

az emberi méltósághoz, a jó hírnévhez és a magántitok védelméhez fűződő alkotmányos jogokba.

A következő problémát az jelenti, ha elhunyt személyek képének (hangjának) felhasználásával kerül sor a deepfake videó (hanganyag, kép) elkészítésére. Az elhunyt személyek adatvédelmi kérdései[100] sajátos megvilágításba kerülnek a deepfake-kel összefüggésben.[101] És itt nemcsak az a probléma merül fel, ha egy olyan videót alkotnak, amelyen egy elhunyt személy szerepel, hanem sokkal komolyabb (és speciálisabb) probléma, ha például egy elhunyt személy stílusában készül egy olyan új lemez, amelyen lévő számokat teljes egészében az MI hozza létre. Önmagában valakinek a stílusában készíteni egy lemezt álláspontom szerint nem jelent sem adatvédelmi, sem szerzői jogi kérdést. Az olyan nagy együttesek, mint például a The Beatles, eddig is számos követőt vonzottak, akik aztán jobb vagy rosszabb minőségben, de a The Beatles stílusában jelentettek meg lemezeket. A probléma ott kezdődik, ha az MI lemásolja a stílust, majd ebben a stílusban John Lennon eredeti hangjával azonos hangon, George Harrison és Paul McCartney gitárstílusában, Ringo Starr dobolásával készít felvételeket. Ki (vagy MI?) ilyenkor a dal szerzője? Kit illet a jogdíj? Ha az MI lesz a jogtulajdonos, feltüntethetik-e, hogy a „The Beatles (MI által készített) legújabb albuma”? Számtalan kérdés merül fel, amire nagyon nehéz jogilag megfelelő választ adni.

Talán kevésbé jelent gondot az MI által kreált személyek jogi szabályozása.[102] Ha az MI nem egy létező személy arcképét, videóját, hangját használja fel a deepfake elkészítéséhez, hanem egy olyan személyt hoz létre, akinek nincs megfelelője a való világban, akkor az legalább azt a jogi problémát nem veti fel, hogy sérülnek-e a létező (akár elhunyt) személy személyiségi jogai, ezzel kapcsolatban nem merülnek fel adatvédelmi problémák sem. Az más kérdés, hogy feltétlen jogi szabályozást igényel az így létrehozott személyek „jogállása”: kit illetnek a vele kapcsolatos jogok, kit illetnek a vele kapcsolatosan befolyó bevételek, megilleti-e a büntetőjogi védelem? Az már egy különösen pikáns helyzetet teremtene, ha az MI által létrehozott „személy” sérelmére követne el valaki más hamisítást, azaz hozna létre egy olyan deepfake videót, amelyen az MI által kreált személy szerepel anélkül, hogy neki erre joga lenne.

A deepfake felveti a közéleti szereplők hatékonyabb védelmének kérdését, valamint a nyilvánvalóan valótlan, hamis nyilatkozatok terjesztésének a kriminalizálását is. Bár az EJEB szerint a közszereplőknek el kell viselniük a magánéletükbe való nagyobb beavatkozást, adott esetben azt is el kell fogadniuk, hogy kigúnyolják őket,[103] ahogyan Sorbán rámutat, a von Hannover kontra Németország ügyben az EJEB azt is kimondta, hogy a közszerepléssel össze nem függő élethelyzetekben a közszereplőt is megilleti a személyes identitásának a védelme (ide tartozik

[100] Az adatvédelemmel összefüggésben utalni kell az Európai Unió és Tanács 2016. április 27-i (EU) 2016/679. sz. általános adatvédelmi rendeletére (GDPR).

[101] Van der Sloot–Wagensveld–Koops: i. m., 9.

[102] Van der Sloot–Wagensveld–Koops: i. m., 9.

[103] Van der Sloot–Wagensveld–Koops: i. m., 11.

a képmás védelme[104] is).[105] Lényegében ezt foglalja össze a Ptk. 2:44. §-ában foglalt magyar jogi szabályozás is.[106]

Ugyanígy külön szabályozást igényelnek a választással kapcsolatosan deepfake segítségével elkövetett cselekmények is, azok kiemelkedő tárgyi súlya miatt (akár a Btk. 350. §-ának megfelelő kiegészítésével).

Nyilvánvalóan azt is jogilag kell szabályozni, hogy milyen intézetek, milyen eszközökkel végezhetik a deepfake detektálást, és ehhez milyen jogkörök járnak.[107] Jó módszernek tűnik, ha a deepfake tartalmakat kötelező ellátni vízjelekkel és egyéb címkékkel (metaadatokkal),[108] de Feeney rámutat arra, hogy e vízjelek és metaadatok eltávolítása is könnyen automatizálható.[109] Az Európai Unió mesterséges intelligenciáról szóló jogszabályának módosítására irányuló javaslat 52. cikk (3) bekezdés 1. albekezdése szerint azon MI-rendszerek felhasználói, amelyek olyan szöveges, audio- vagy vizuális tartalmat generálnak vagy manipulálnak, amely megtévesztő módon eredetinek vagy valóságosnak tűnhet, és amely olyan személyeket ábrázol a beleegyezésük nélkül, akik látszólag olyan dolgokat mondanak vagy tesznek, amelyeket nem mondtak vagy tettek („deepfake"), kellő időben, megfelelő, egyértelmű és látható módon közlik, hogy a tartalmat mesterségesen hozták létre vagy manipulálták, valamint – amikor csak lehetséges – közlik a tartalmat generáló vagy manipuláló személy nevét is. A közlés a tartalom olyan, a tartalom címzettje számára jól látható címkézése formájában valósul meg, amely jelzi, hogy a tartalom nem eredeti. A tartalom címkézésekor a felhasználók figyelembe veszik a technika általánosan elismert állását, valamint a vonatkozó harmonizált szabványokat és előírásokat.

A deepfake-kel kapcsolatosan

– tiltani kell a deepfake tartalmak nem saját célra történő előállítását,
– az ezekhez alkalmas technológia ilyen célra történő árusítását és
– a deepfake tartalmak megszerzését, birtoklását, valamint
– korlátozni kell a tárhelyszolgáltatókat (akár előzetes ellenőrzési kötelezettség előírásával).

[104] Michael Jordan NBA legenda 2015-ben 8,9 millió dollárt kapott pusztán azért, mert a Jewel Food Stores áruházlánc az engedélye nélkül használta fel az arcát. https://www.espn.com/nba/story/_/id/13486052/supermarket-chain-pay-michael-jordan-89-million-use-name (2023. 07. 07.).

[105] Sorbán: i. m., 87.

[106] *„Ptk. 2:44. § (1) A közügyek szabad vitatását biztosító alapjogok gyakorlása a közéleti szereplő személyiségi jogainak védelmét szükséges és arányos mértékben, az emberi méltóság sérelme nélkül korlátozhatja; azonban az nem járhat a magán- és családi életének, valamint otthonának sérelmével. (2) A közéleti szereplőt a közügyek szabad vitatásának körén kívül eső közléssel vagy magatartással szemben a nem közéleti szereplővel azonos védelem illeti meg. (3) Nem minősül közügynek a közéleti szereplő magán- vagy családi életével kapcsolatos tevékenység, illetve adat."*

[107] Van der Sloot–Wagensveld–Koops: i. m., 10–16.

[108] Az Amerikai Egyesült Államokban egy 2019-es (végül meg nem szavazott) szövetségi törvényjavaslat (Deepfakes Accountability Act) már tartalmazta azt a rendelkezést, hogy minden deepfake tartalmat vízjelekkel kell ellátni. Delfino: i. m., 303.

[109] Feeney: i. m., 9.

Emellett fontos szerep jut a figyelemfelhívó kampányoknak és a deepfake áldozatai megsegítésének is. Utóbbi szervezetek között emeli ki Brooks a Cyber Civil Rights Initiative, az EndTab, a National Suicide Prevention Lifeline, a Cybersmile, az identitytheft.gov, a withoutmyconsent.org, a Google súgó és az Imatag szerepét.[110]

4. Összegzés

A deepfake tartalmak elterjedésének a veszélye napjaikban egyre nagyobb. Ahogyan arra Mráz is helyesen rámutat,[111] ennek két oka van: egyrészt egyre több „nyersanyag" kerül az internetre, azaz szinte mindenkiről készülnek olyan kép- és videófelvételek, amelyek az interneten megtalálhatóak, ezáltal a deepfake tartalmak elkészítését elősegíthetik. Másrészt a technológia rohamosan fejlődik, így egyre kevesebb ilyen anyagra van szükség ahhoz, hogy szinte (?) tökéletes hamisított tartalmakat hozzanak létre. Márpedig az emberek jobban hisznek a szemüknek, mint bármi másnak (*„seeing-is-believing heuristic"*[112]). Erre mutat rá Köbisék felmérése, amelyből négy fontos következtetést vontak le:

1. az emberek egyre kevésbé képesek észlelni a hamisításokat;
2. ezt az észlelési pontosságot sem figyelemfelkeltéssel, sem anyagi ösztönzőkkel nem lehet növelni (azaz az észlelési hibák inkább a képtelenségből, mintsem a motiváció hiányából erednek);
3. az emberek sokkal inkább hajlamosak arra, hogy a hamis videókat valódinak tekintsék, mint arra, hogy a valódi videóra azt mondják, hogy hamis, és emellett
4. messzemenően túlbecsülik saját észlelési képességüket.[113]

Nem csoda, hogy a Pentagon már 2018-ban több tízmillió dollárt költött a deepfakekel kapcsolatos kutatásra és a rosszindulatú deepfake elleni harcra.[114] A deepfake elterjedésének az egyik legfőbb veszélye, hogy alááshatja a videó-, kép- és hangfelvételek valódiságába vetett általános bizalmat,[115] különösen azáltal, ha az eredeti videók és a hamisítások egymástól megkülönböztethetetlenné válnak.[116] Ugyanúgy, ahogy az álhírek (fake news) megkérdőjelezik a valós hírek hitelességét,[117] a deepfake is megkérdőjelezi a valós tartalmak valódiságát.

De akármilyen jó is lenne a deepfake szabályozása, várhatóan nem zárná ki, hogy azt megkerüljék. Ez azonban nem szabad, hogy elvegye a kedvét a jogalko-

[110] https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf (2023. 07. 07.).

[111] Mráz: i. m., 251.

[112] Frenda, Steven J.–Knowles, Eric D.–Saletan, William–Loftus, Elizabeth F.: False memories of fabricated political events. *Journal of Experimental Social Psychology,* 2013/2. 280–286.

[113] Köbis, Nils C.–Doležalova, Barbora–Soraperra, Ivan: Fooled twice: People cannot detect deepfakes but think they can. *iScience,* 2021/11. 1., 7.

[114] https://futurism.com/the-byte/pentagon-ai-director-deepfake-protections (2023. 07. 07.).

[115] Palmiotto: i. m., 1.

[116] Mosley, Tonya: Perfect Deepfake Tech Could Arrive Sooner Than Expected. *Wbur,* October 02, 2019. https://www.wbur.org/hereandnow/2019/10/02/deepfake-technology (2023. 07. 07.).

[117] Ürmösné–Nyitrai: i. m., 87.

tónak. Ahogyan az USA-ban sem lehet egyetérteni a nihilistákkal, akik szerint, ha a bűnözők nem követik a fegyvertörvényt,[118] akkor nincs is szükség arra, ugyanúgy nem szabad meghátrálni pusztán azért, mert egyelőre nem lehet megjósolni, hogy a deepfake jogi szabályozása mennyire lehet hatékony. Ennek megfelelően az összes, a tanulmányban érintett jogágban (közjogban, polgári jogban, büntetőjogban) mielőbb jogi szabályozási lépésekre van szükség. Bár a tanulmány bemutat néhány olyan, jelenleg is érvényben lévő jogszabályi rendelkezést, amely érintheti a deepfake területét, ezekkel nem szabad megelégedni. Továbbá ehhez nem elegendő a nemzetállami szabályozás, hanem (ahogyan az Európai Uniós szabályozási tervezet is mutatja) nemzetközi szinten is érinteni kell a kérdést, mivel a deepfake a számítógépes környezet jellegéből fakadóan határokon átnyúló problémákat okozhat.

[118] https://www.nraila.org/articles/20190628/study-reinforces-what-we-already-know-criminals-don-t-follow-the-law (2023. 07. 07.).