KözigazgatásTudomány 2024/2. 18–38.
DOI: 10.54200/kt.v4i2.89

HAMIS VALÓSÁG, VALÓS VESZÉLY: DEEPFAKE AZ AI ACT-BEN[1]

Badinszky Áron*

* PhD hallgató, Eötvös Loránd Tudományegyetem, Állam- és Jogtudományi Kar. E-mail: dr.badinszkyaron@gmail.com

Absztrakt

Napjainkban a mesterséges intelligencia számtalan manifesztációja közül valószínűleg a deepfake technológiában rejlik a legtöbb potenciál arra, hogy rövid időn belül látványosan felforgassa világunkat. A deepfake jó kezekben ajándék, hiszen hatékonyabbá teszi a gyógyszerkísérleteket, segítségével „életre kelthetünk" régen elköltözött színészeket, ugyanakkor elsőszámú fegyvere lehet a dezinformációnak, szexuális visszaéléseknek, illetve a kiberbűnözés eszköztárát is számos új elemmel „gazdagította". Az egyre közismertebb aggályok ellenére a közelmúltban elfogadott európai MI szabályozás, az AI Act mégis kevésbé tartja kockázatosnak a deepfake-et, mint például a hallgatói csalásokat szűrő egyetemi rendszereket, hiszen még ez utóbbiakra is jóval szigorúbb követelményeket fogalmaz meg. Alábbi elemzésemben részletesen áttekintem a deepfake jelentette veszélyeket, érzékeltetem azok súlyát, majd az európai szabályozás pro és kontra elemzésével, azt nemzetközi példákon át kontextusba helyezve arra keresem a választ, miért nem szigorúbb, a kockázatok szintjét valóban tükröző rezsimet alakított ki az Európai Unió.

Kulcsszavak

mesterséges intelligencia, AI Act, deepfake, MI szabályozás

Abstract

Among the myriad manifestations of artificial intelligence today, deepfake technology probably has the greatest potential to spectacularly transform our world in a short time. Deepfake is a gift in the right hands: it can make drug research more effective, bring passed actors back to life, however it can be a weapon of disinformation, sexual abuse and added many new elements to the cybercrime toolbox. Despite the growing public concerns, the recently adopted European AI regulation, the AI Act, considers deepfakes to be less risky than, for example, university fraud filtering systems, as even the latter are subject to much stricter requirements. In my analysis below, I will provide a detailed overview of the risks posed by deepfake, give a sense of their severity, and then, by analyzing the pros and cons of European regulation and contextualizing it with international examples, seek to answer the question: why the European Union has not developed a more stringent regime that truly reflects the level of risks posed by deepfake.

[1] A jelen tanulmány a K-142232 OTKA_22 alapkutatási pályázat keretében született a KIM és az NKFIH támogatásával.

Keywords

artificial intelligence, AI Act, deepfake, AI regulation

1. Képbehozó

Idén nyár közepén mérföldkőhöz érkezett a globális mesterséges intelligencia (MI) szabályozás, ugyanis megszületett a köztudatba AI Act-ként[2] bevonult rendelet (a továbbiakban: AIA vagy Rendelet), ami a világon az első, MI-t átfogó jelleggel, nemzetközi szinten rendező jogszabály (Council of the EU, 2024). A Rendelet sokáig formálódott a digitalizációs robbanást követő jogalkotási „cunamiban", tartalma számottevően bővült, ami elsősorban a technológia szüntelen fejlődésének[3] köszönhető. Az AIA fogadtatása alapvetően pozitív volt, de mind az akadémia (Hacker, 2023), mind a piac irányából kapott kritikát, elsősorban azokkal a rendelkezéseivel kapcsolatban, amelyek már ma idejétmúltnak tűnnek a technológia fejlettségének tükrében. Véleményem szerint az AIA-nek gyermekbetegségei mellett van egy igen komoly – annak veszélyes természete ellenére sem hangsúlyozott – hiányossága, amelyre az alábbiakban kívánom felhívni a figyelmet.

A jogalkotás – így az európai reguláció is – mindig értékválasztással jár. Az Európai Bizottság kezdetektől az emberközpontú[4] – azaz az EU Alapjogi Chartában és alapszerződéseiben biztosított alapjogokat végsőkig szem előtt tartó és érvényesítő[5] – rezsim kialakítását tűzte ki célul. A jogalkotó dolgát aligha könnyítette meg az MI-t körülvevő óriási „globális zaj", amikor a választott értékek védelmét valóban biztosító szabályrendszer megalkotásán munkálkodott. Talán így eshetett, hogy egy kérdésben biztosan alábecsülte az MI jelentette veszélyeket, és elégtelenül határozta meg az alapjogok védelméhez szükséges a garanciák szintjét. A kritikus pont az ún. deepfake tartalmak létrehozását lehetővé tévő MI rendszerek alacsony kockázati kategóriába való besorolása volt, aminek következtében a deepfake-el szembeni jogi fellépés szintje elmarad a technológia által a Chartában rögzített alapjogokra és azok érvényesítésére jelentett veszélyektől. Jelen tanulmány célja annak bemutatása, hogy a deepfake nemcsak egy a sok MI alapú technológia közül, hanem ártó kezekben olyan veszélyt jelent, aminek kockázatait véleményem szerint az AIA jelenlegi formájában nem kezeli kielégítően. Ennek megalapozása érdekében bemutatom a deepfake természetét, főbb felhasználási módjait, áttekintem az AIA szabályozási struktúráját, válaszokat keresve arra, hogy a kódex rendszerében miért ott helyezkedik el a deepfake, ahol. Célom rávilágítani arra, hogy a technológia által támasztott veszélyek fényében miért tartom hibás – és mielőbb kezelendő – döntésnek a jogszabály rendelkezéseit.

---

[2] 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).

[3] Az ún. általános célú MI-modellek – angolul *foundation model* vagy *general-purpose AI* – például nem szerepeltek az EU Bizottság által 2021 áprilisában előterjesztett anyagban. Idővel azonban – különösen az ilyen technológián alapuló ChatGPT 2022 novemberi berobbanását követően – a terület reflektorfénybe került, szabályozása fokozatosan kikristályosodott és beépült a Rendeletbe – lásd: Rendelet V. fejezet.

[4] Az alapelvvel kapcsolatban lásd: A mesterséges intelligenciáról szóló összehangolt terv COM(2018) 795 final 1. oldalát.

[5] Lásd: Európai Bizottság (2019, 12).

2. A deepfake jelleme

A „deep learning”[6] és a „fake” szavak összeragasztásából kreált elnevezés egyaránt vonatkozik magára az MI alapú megoldásra, és az ennek felhasználásával létrehozott eredményre. Definícióját több jellegadó elem körvonalazza: alapvetően olyan technológia, amivel digitálisan manipulált, hiperrealisztikus képek, videók, hangfelvételek készíthetők (Davies, 2024; Kan, 2019; Zinski, 2020), amelyeken a szereplők olyan dolgokat tesznek vagy mondanak, amelyek valójában meg sem történtek (Westerlund, 2019, 40). Jellemzően emberi hang, arcok kicserélésére, módosítására használják, de más célokra is alkalmazható (Chadha et al., 2021, 558–560; Europol, 2024, 9). A deepfake szintetikus – azaz MI segítségével kreált – média (Europol, 2024, 5; Payne, 2024), amely valamilyen létező médiatartalom átalakításával, vagy akár teljesen „eredeti”, tetszőleges tartalmú anyag létrehozásával keletkezik. A technológia nem tekint vissza hosszú múltra: ugyan a kezdetleges arccserélő appok már 2013-tól (Dredge, 2016) elterjedtek az okostelefonok használói közt, de az első, hivatalosan is deepfake-nek nevezett videók 2017-ben robbantak be a köztudatba.[7] A deepfake a mikrochip alapú univerzum szülötte, s létezése is elválaszthatatlan a digitális médiától. Készítése számítógéphez, terjesztése pedig elektronikus médiumokhoz kötött, így az ilyen tartalmak elsősorban az interneten – jórészt a közösségi médiában (Weimann & Masri, 2023, 5) – keringenek, de előfordult, hogy a televízión vagy akár telefonhívásokon (Damiani, 2019; Somers, 2020) keresztül is találkozhattak vele.

A technológia oldaláról megközelítve a deepfake előállítása kezdetben több mélytanuló[8] algoritmus összehangolt munkájának eredménye volt (Westerlund, 2019, 41). Mára azonban túlnyomó többségüket az MI egyik – jelenleg is beláthatatlan potenciállal rendelkező – fejlődési ágazatát képező, ún. generatív mesterséges intelligencia[9] segítségével hozzák létre (Davies, 2024). E technológia új távlatokat nyitott (Hurst, 2023) a deepfake-ek készítésben, hiszen míg korábban csak valóban létező médiaanyagból tudtak dolgozni az alkotók – értsd: csak valós szituációban rögzített valódi ember arcát, hangját tudták módosítani, pl. kicserélni egy másik, valóban létező emberével –, a generatív MI segítségével bármilyen tetszőleges élethelyzet, környezet, történés könnyedén létrehozható, modellezhető. Utóbbira kiváló példa az, a világsajtót bejárt kép, amikor Ferenc pápa fehér pufi dzsekiben látható (Huang, 2023), de biztos, hogy már sokan találkoztak olyan – ugyan még az „előző generációs” technológiával készített – videóval, ahol ikonikus filmek jeleneteiben az eredeti szereplők arcát egy másik világsztáréra cserélték (Holliday, 2021, 9–10). A deepfake létéhez tehát mindenképpen valamilyen MI alapú megoldás szükséges, azonban míg korábban az előállítható képek, videók tartalmát leszűkítette a valóság, a generatív MI megjelenésével szó szerint csak a képzelet szab határt az alkotásnak. Alább tárgyaljuk, hogy e képzelet sajnos többnyire rosszat akar.

---

[6] A deep learning a gépi tanulás egyik formája, amely az emberi agy működését modellező MI technológián, úgynevezett neurális hálókon alapul. Lásd: Holdsworth & Scapicchio (2024).

[7] Ekkor egy máig ismeretlen Reddit felhasználó feltöltött néhány, hollywoodi színésznőket szexuálisan explicit helyzetben „ábrázoló” videót a közösségi oldalra. Ő használta először a deepfake kifejezést is. Lásd: Ajder et al. (2019, 3); Maddocks (2020, 415).

[8] A gépi tanulás – *machine learning* – egyik ágazata a *deep learning.* Lásd: Holdsworth & Scapicchio (2024).

[9] A generatív MI lényege, hogy az algoritmus a felhasználó által megadott utasítások alapján szöveget, hangot képet vagy mozgóképet állít elő. Ilyen technológián alapszik a közismert ChatGPT, Dall-E vagy a Stable Diffusion alkalmazás, amelyek mellett egy sor másik, hasonló célú, akár ingyen hozzáférhető megoldás is létezik.

Az említetthez hasonló esetek ma gombamód szaporodnak, hisz napjainkra bárki hobbiból belefoghat a „házi”[10] deepfake gyártásba (Ajder et al., 2019, 4; Westerlund, 2019, 41). A jelentős demokratizálódás alapvetően három tényezőnek tudható be: az első, hogy a számítástechnika forradalma következtében a deepfake előállításához szükséges eszközök és szoftveres alkalmazások olcsóbbá, így széles körben elérhetővé váltak (Hurst, 2023). Tovább segíti a helyzetet, hogy leomlottak a kompetencia-hiányból fakadó akadályok. Ma már nem szükséges komolyabb számítógépes tudás a deepfake létrehozásához, hiszen a még legmodernebb MI is egyszerűen hozzáférhető: némi internetes utánajárással, YouTube videók tanulmányozásával bárki (Ajder et al., 2019) „alkotóvá” avanzsálhat. Kezdetben a deepfake tartalmak generálásához komoly informatikai ismeretekre és több száz képre volt szükség az érintett személyről, míg napjainkra elég egy okostelefon (Kelleher, 2023) és mindössze egy rövid videó, kép vagy néhány másodperc hang (Jain, 2023; Kan, 2019) az adott illetőről, hogy valósághűen[11] „reprodukálni” lehessen. Meghökkentő adat, de ma kevesebb, mint 25 perc munkára és egyetlen jó minőségű képre van szükség ahhoz, hogy valakiről egy 60 másodperces pornográf deepfake videót generáljanak (Security Hero, 2023), ráadásul mindez nem kerül semmibe, hiszen számos,[12] e célra ingyenesen elérhető MI szoftver létezik. A deepfake-hez szükséges „alapanyag” ráadásul könnyedén hozzáférhető a közösségi médiában uralkodó megosztási és kitárulkozási kultúrának köszönhetően. Az utolsó, de talán leginkább meghatározó ok az internet jelentette névtelenség és biztonság, amely a világháló sajátosságaiból, illetve a különböző rejtőzködésre tervezett eszközök – szoftverek és a technikai megoldások – fejlettségéből és hozzáférhetőségéből fakad. Ezt kihasználva sokan veszik a bátorságot arra, hogy vicces, botrányos, de akár bűncselekménynek is minősülő deepfake tartalmat állítsanak elő. Hatóság legyen a talpán, aki a mai világban meg tudja találni azt az elkövetőt, aki egy kicsit is figyel arra, hogy elrejtőzzön.

3. Szép új világ?

A technológiával kapcsolatban régi igazság, hogy erkölcsi minőségét felhasználóitól kapja: önmagában nem lehet jó vagy rossz, azzá mindig az ember teszi – a fegyverekkel például egyaránt meg tudjuk védeni a törvényt, de át is tudjuk hágni azt. Nincs ez másképp a deepfake esetében sem. A technológia – pozitív hozadékai mellett – a kiberbűnözés, különösen a szexuális jellegű, közrend- és vagyon elleni bűncselekmények eddig ismeretlen formáit tette lehetővé. A deepfake-visszaélések nem csak egy elszigetelt, vagyonos és szuperintelligens bűnözői elit számára kivitelezhetők, hanem tömegjelenséggé váltak. A következőkben lényegre törően bemutatom azt a két területet, amelyeket a leginkább átsző és megmérgez a rosszindulatú deepfake-ek terjedése, hogy érzékeltessem: igen ellentmondásos, káros és veszélyes technológia vált elérhetővé a fejlett világ számára. Ezek fényében lássuk a megdöbbentő, elgondolkodtató, és talán kissé baljós helyzetet:

[10] A deepfake tartalmak legnagyobb hányadát jelenleg „amatőr” egyének vagy kisebb csoportok állítják elő.

[11] Sokáig az ilyen tartalmak minősége is hagyott maga után kívánnivalót, így könnyen felismerhető volt. Ezeket nevezték *shallow fake-nek* – lásd: Ajder et al. (2019, 11) – vagy *cheap-fake*-nek. Napjainkra azonban az MI az ember számára felismerhetetlenül hiteles tartalmakat képes gyártani.

[12] Lásd: Free Deepfake, Online: https://theresanaiforthat.com/s/free+deepfake/

3.1. Tömeges szexuális visszaélés

Az interneten keringő összes deepfake videó 96–98%-a pornográf jellegű (Ajder et al., 2019, 1; Security Hero, 2023). Elérésükhöz még csak nem is kell elmerülni a dark-weben (Mahdawi, 2023), bárki könnyedén megtalálja, aki szeretné. Az explicit tartalmak elárasztják az internetet: 2023-ban csaknem százezer (Security Hero, 2023) ilyen videó volt megtalálható online, ami több mint hatszorosa a 2019-es adatoknak (Ajder et al., 2019, 1). Ezekben szinte kizárólag olyan személyek szerepelnek, akik ebbe egyáltalán nem egyeztek bele (Lawelle, 2024; Naezer & van Oosterhout, 2021, 80–81), ami nem meglepő a megalázó, nem egyszer erőszakos és kifejezetten megszégyenítő eredmény tekintetében. Ha a hamisított, non-konszenzuális tartalmak áradata nem volna elég, korunkban egyre inkább terjed a deepfake bosszúpornó, amikor elvetemült emberek korábbi partnerükről készítenek explicit videót anélkül, hogy ténylegesen lefilmeznék az áldozatot (Kelleher, 2023; Maddocks, 2020, 417; Westerlund, 2019, 37). Az anyagot vagy az érintett zsarolására használják, vagy egyszerűen elérhetővé teszik a nyilvánosság számára. Szomorú tendencia, hogy a pornográf deepfake-ek szinte száz százaléka – akár kiskorú (Hao, 2020b) – nőket ábrázol (Ajder et al., 2019, 2; Hao, 2021; Security Hero, 2023). Közöttük két érintetti kört lehet elkülöníteni, akiknek emberi méltósága, személyiségi jogai és lelki-mentális egészsége is kétségtelenül súlyosan sérül az ilyen videók publikálását követően. Az egyik csoportot azok a színésznők (Kelleher, 2023), influencerek (Mahdawi, 2023) vagy más közszereplők, így akár politikusok (Németh, 2024; US Senate, 2024) alkotják, akik szerepeltetése ilyen kontextusban mára sajnos általánossá vált,[13] míg a másik – nagyságrendekkel kisebb – halmazba azok a civilek tartoznak, akikről jellemzően rosszakaróik készítenek deepfake-pornót.

A deepfake pornó létezése és az azokban akaratuk ellenére szereplők jogainak sokszoros és súlyos megsértése már önmagában is komoly probléma, azonban a helyzetet még sötétebbé teszik áldozatok életét egy-egy ilyen tartalom nyilvánosságra kerülésekor érő, olykor tragikus következmények. A szexuális tartalmú deepfake videókban való feltünést ugyanis szinte lehetetlen megakadályozni, az áldozatok pedig gyakorlatilag eszköztelenek a támadóval szemben (Tenbarge & Chan, 2023). A valóság bizonyítása az esetek nagy részében már okafogyott, hiszen a „szellem kiszabadult a palackból". Ahogyan arra számos külföldi (Ryan-Mosley, 2023) sőt, hazai (RTL, 2024) precedens is van, a szexuálisan explicit deepfake-ek egyre jobban terjednek a fiatalok körében, sajnos nem ritka, hogy maga az elkövető is fiatalkorú. Elkeserítő az is, hogy a deepfake pornó gyakori eszközévé vált (Narvali et al., 2023) a cyberbullying-nak,[14] ami sajnos drámába is fordulhat: a közelmúltban egy 14 éves lány vetett véget életének, amikor iskolájában elterjedtek a róla készített, hamisított képek (Linn, 2024). Nem kérdés: a könnyű hozzáférhetőség és a kellő védelmi mechanizmusok hiánya kiegészülve a gátlástalansággal komoly veszélyt jelentenek az emberi jogokra.

3.2. A valóság válsága

Koltay (2020, 303) szerint: „A technológiai fejlődés a hazugságok újabb és újabb generációjának megszületését támogatja", ami nem is lehetne igazabb. A deepfake segítségével kedvünk

[13] A legfrissebb kutatások alapján a szexuális tartalmú deepfake-ek 94%-a a szórakoztatóiparban tevékenykedő személyeket ábrázol. Lásd: Security Hero (2023).

[14] A szakirodalom összefoglalva így utal a digitális térben és eszközökkel, jellemzően az interneten és a közösségi médiában zajló zaklatásra. Lásd: UNICEF (2024).

szerint alakíthatjuk a valóságot, ami egy olyan világot eredményez(ett), ahol megrendült a médiába, az információkba, de sajnos a valóságba vetett bizalom is. Az ember már a szemének sem hihet: az egyre kifinomultabb MI egyre meggyőzőbb és nehezebben kimutatható szintetikus média létrehozását teszi lehetővé (Sensity, 2024, 5), amit számtalan, akár társadalmi szintű kockázatot eredményez.

A történelem megmutatta, hogy a politikai hatalom megszerzésében érdekeltek soha nem rettentek vissza a szavazók manipulálásától, az MI pedig ebben is új perspektívákat nyitott. 2024 valóságos választási „szuperév" volt, hiszen a Föld lakosságának csaknem fele járult az urnákhoz (Ewe, 2023; Guterman, 2024). A klasszikus lejáró kampányok idejének vége, hiszen világszerte a valóságtorzítás korát éljük, aminek kulcseszköze a deepfake. A Stanford jogászprofesszora szerint: „Az MI és a demokrácia kapcsolatának alapszabálya az, hogy [az MI] felerősíti a rendszer minden jó és rossz szereplőjének képességeit, hogy elérjék ugyanazokat a céljaikat" (Guterman, 2024). A 2020-ban a Delhiben tartott választások során a kormányzó párt elnöke, Manoj Tiwari deepfake segítségével törte át a nyelvi akadályokat, amelyek elválasztották számos potenciális választójától (Bodnár, 2020). Nem csoda, hogy az amerikai elnökválasztásba is beszivárgott az MI (De Luce, 2024). Mindkét oldal – illetve névtelenségbe burkolódzó támogatóik[15] – előszeretettel vetette be a technológiát, a saját népszerűségének növelése (Molnár, 2024; Spring, 2024) és a másik – akár azonos pártba tartozó – jelölt lejáratása érdekében (Devlin & Cheetham, 2023; Elliott, 2024). Ennek tankönyvi példája az, a Joe Biden elnök megtévesztésig tökéletes hangján megszólaló, hamisított automata telefonhívás volt (Ahmed, 2024), amiben „arra kérte" a new hampshire-i szavazókat, ne menjenek el az előválasztásra, hanem őrizzék meg szavazataikat novemberig. A hívást Biden legfőbb demokrata kihívójának tanácsadója rendelte meg mindössze 500 dollárért (Shepardson, 2024a), ám végül visszafelé sült el a dolog, ugyanis a lebukást követően a hívást közvetítő cég egymillió (Feiner, 2024), míg a tanácsadó hatmillió dolláros bírságban részesült (Shepardson, 2024b). Az USA-ban már egyenesen attól tartottak, hogy nemcsak a választások előtt, de utána is komoly felfordulást okozhat az elszabaduló, deepfake-re építő dezinformációs hullám, amely a választások állítólagos elcsalását, manipulálását terjeszti (Taylor, 2024). A félelem ugyan alaptalannak bizonyult (Carpenter, 2024), de a közösségi médiában már a választás másnapján elkezdtek terjedni olyan deepfake videók, amelyen az állítólagos Joe Biden pikírt stílusban kritizálja az alulmaradó demokrata elnökjelöltet és Donald Trumpot is.

Dezinformáció nem csak a politikai véleményformálásban zajlik, hanem a hadszíntéren is, a deepfake a kiberhadviselés[16] egyik pusztító eszközévé lett, a dezinformáció pedig elképesztő erejű fegyver. A közelmúltban történt események – pl. orosz-ukrán háború, izraeli terrorcselekmények (Sensity, 2024; Wakefield, 2022) – nyomán is számos olyan digitális tartalom járta be az internetet, amelynek semmilyen valóságalapja nem volt, azonban első ránézésre teljesen hitelesnek tűnhettek. Céljuk az igazság tökéletesen megtévesztő eltorzítása, ezzel a hátország lakóinak megtévesztése, pánikkeltés és a nemzetközi közvélemény félrevezetése volt. A dezinformációval nem csak a világpolitikát, hanem az ember mikrokörnyezetét is lehet befolyásolni, aminek igen tanulságos esetei, amikor egy buzgó anyuka lejáratta a lányáéval rivális cheerleader csapatot (Vella, 2021), vagy amikor egy iskolaigazgatót majdnem elbocsátottak az állásából, mert egyik kollégája bosszúból olyan deepfake-el manipulált hangfelvételeket terjesztett, „amelyeken főnöke rasszista és antiszemita kirohanásokat tesz" (Bitport, 2024).

[15] Nem egy esetben felmerült, hogy az ilyen tartalmakat a választások kimenetelét befolyásolni akaró államok által támogatott hackercsoportok készítik és terjesztik. Lásd: Trish (2024); Ventura (2024).

[16] Lásd a Sensity nagyszerű tanulmányát a deepfake és a kiberhadviselés kapcsolatáról (Sensity, 2024).

A deepfake hozzáférhetősége következtében szintén népszerű az emberi hang hamisítása, ami melegágya az az újgenerációs telefonos csalásoknak is. A hanghordozást, intonációt, beszédhibákat is tökéletesen reprodukáló technológiának nemcsak az amerikai elnökválasztás során estek áldozatául a gyanútlan fülek. Az első feljegyzett ügyben több mint kétszázezer eurót csaltak ki illetéktelenek egy nemzetközi cég vezetőjétől, aki azt hitte, német főnökével beszél, aki arra utasította, utalja át az összeget egy magyar bankszámlára (Damiani, 2019). Mivel még a csalók által generált német akcentust is hitelesnek tartotta, gondolkodás nélkül átutalta a pénzt, aminek természetesen nyoma veszett. Gondoljunk bele, milyen ajtókat nyithat a technológia terjedése például az „unokázós csalók" számára. A módszert következő szintre emelték azok a csalók, akik egy hong-kong-i cég pénzügyi alkalmazottjának egész deepfake-videókonferenciát szerveztek, ahol a jóhiszemű kolléga a „megjelent" vezetői utasítására több mint 25 millió dollárt utalt a bűnözőknek (Chen & Magramo, 2024).

Az ártó gyakorlatokat hosszan sorolhatnánk, ám úgy vélem, a kockázatok világosak. Az MI folyamatos fejlődésben van, így valószínű, hogy még korántsem merítették ki a deepfake nyújtotta ártó lehetőségek tárházát. Az viszont már most nyilvánvaló, hogy a technológia sokrétű veszélyeket hordoz, amelyek mérsékléséhez átfogó, széles eszköztárat felvonultató megközelítés szükséges, aminek egyik eleme a hatékony jogi szabályozás lehet(ne).

3.3. Az érem másik oldala

Noha a helyzet igen megrázó, igazságtalan volna a deepfake-ről kizárólag negatív kontextusban beszélni, hiszen kétélű fegyverrel van dolgunk, amelynek számos, kifejezetten hasznos, etikus felhasználási módja, területe van. A technológia legnagyobb haszonélvezője egyértelműen a szórakoztatóipar: a deepfake megjelenésével olyan lehetőségek tárultak a filmszakma elé, melyek önmagukban is sci-fi-be illenek, hiszen rég elköltözött színészeket kelthetnek életre, és akár olyan produkciókban is castingolhatják őket, amelyeket már haláluk után készítettek (Cheng, 2013; Hao, 2020a). Az elmúlt években láthattunk olyan alkotást is, amelyben ma már korosabb világsztárok deepfake-el megfiatalítva tűntek fel a vásznon,[17] de az MI új távlatokat nyitott az utómunkában is, nagyságrendekkel csökkentve a produkciós költségeket (Hood, 2021). A technológia nemcsak Hollywoodba tört be, néhány éve az avantgárd egyik leghíresebb képviselőjét, Salvador Dalít keltették életre deepfake segítségével a művész halálának harmincadik évfordulóján (The Dalí Museum, 2019). Az emberi hang klónozására képes deepfake megoldások forradalmasították a nemzetközi környezetben való munkavégzést,[18] megnyitották az eddig nyelvi korlátok miatt elérhetetlen tömegek számára is a kultúrában, politikában való részvételt, valamint hozzáférést teremtettek különböző edukatív tartalmakhoz is (Westerlund, 2019, 41). A legnagyobb potenciál pedig talán az egészségügyi felhasználásban van: „a deepfake komoly segítséget nyújthat sérült, beteg embereknek egy jobb életminőség eléréséhez. Így a bénulás bizonyos formáiban, pl. ALS-ben […] szenvedők a saját hangjukon tudnak beszélni" (Herke, 2023, 7).

[17] Így nyerte vissza régi formáját például Samuel L. Jackson a Marvel Kapitányban, vagy a Leia hercegnőt alakító Carrie Fisher a Star Wars Rogue One-ban. Lásd: Black & Fullerton (2020); Tsui (2022); Winick (2018).

[18] Ma már nem csak arra képes a technológia, hogy tökéletesen reprodukálja egy személy hangját, hanem arra is, hogy az adott illető ezen a hangon, de más nyelven és akár valós időben szólaljon meg. Ld. Speechify (2024).

4. Deepefake az AIA-ben

4.1. A Rendelet alapelvei

Az Európai Unió 2018-tól kezdve számos dokumentumban[19] körvonalazta azokat az elveket, alapértékeket, amelyek mentén meghatározta az öreg kontinens számára kívánatos MI szabályozás metodikáját és etikai sarokköveit. A mesterséges intelligencia európai megközelítésének AIA-ben kikristályosodott rendszere három pilléren nyugszik: emberközpontú, technológiasemleges, kockázatalapú. Az elvek tartalmát a jogalkotó részletszabályok formájában számos ponton artikulálta, konkretizálta a Rendeletben. E tanulmány szempontjából a legfontosabb pillér az emberközpontúság, aminek vezérelve, hogy az európai MI szabályozás az emberért van, az embert védi, és a lehető legteljesebb mértékben juttatja érvényre az emberi jogokat. Utóbbiakat az EU Alapjogi Chartája rögzíti, de természetesen visszatükröződnek az Alapszerződésekben, valamint a legtöbb közösségi jogszabály is origóként tekint rájuk, amikor célját, etikai kereteit rögzíti. Az Európai Bizottság már jóval az AIA benyújtását megelőzően[20] zászlajára tűzte, hogy emberközpontú MI szabályozási instrumentumot hoz létre. E törekvéseit később maga a jogszabály szövege is több helyütt rögzítette.[21] Sajnálatos módon éppen az emberközpontúság elvének maradéktalan megvalósulását teszik kétségessé a Rendelet deepfake-re vonatkozó szabályai.

A technológiasemleges szabályozási metodikának köszönhetően a lehető legtöbb MI alapú rendszer az AIA hatálya alá tartozik független attól, milyen technológiai megoldás van a „motorháztető alatt". Ennek köszönhetően a Rendelet időtállóbb, hiszen meglehetősen tág és rugalmas MI definíciókkal operál, amely a remények szerint nincs annyira kitéve az exponenciális fejlődés okozta gyors elavulás veszélyeinek. Legyen szó akár mélytanuló algoritmusokról, neurális hálókról, vagy a legmodernebb nagy nyelvi modellekről, ha az MI-rendszer,[22] illetve az úgynevezett általános célú MI modell [23] definíciójának megfelelnek, akkor kiterjed rájuk az AIA hatálya.

A kockázatalapú megközelítés[24] szerint a jogi fellépés szigorúságát az adott MI rendszer vagy modell által az emberekre vagy társadalomra jelentett potenciális veszély szintjéhez igazítják, így „nem lövünk ágyúval verébre". Nem mindegy ugyanis, hogy egy-egy MI alkalmazás milyen mértékben képes befolyásolni a működése által érintett emberek életét, jogait, a közösség biztonságát: egy diszfunkcionális banki hitelbíráló rendszer jelentősebb károkat okozhat az érintettek

[19] Így például: A mesterséges intelligenciáról szóló összehangolt terv COM(2018) 795 final; Fehér Könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése - COM(2020) 65 final; Mesterséges intelligencia Európa számára - COM(2018) 237 final/2.

[20] Uo.

[21] Lásd pl.: AIA preambulum (1)–(2) bekezdés.

[22] AIA 3. cikk (1) bekezdés.

[23] AIA 3. cikk (63) bekezdés – Az angol szövegben *general purpose AI*-nak nevezett megoldások jelentik jelenleg az MI fejlesztés talán legizgalmasabb ágazatát. Olyan algoritmus beszélünk, amelyek működésük során nem feladat-specifikusak, azaz nemcsak az emberi intelligencia egy részterületét – pl. látás – képesek reprodukálni, hanem alkalmazási területek széles palettáján mozognak, mindezt nagyfokú önállóság mellett. Mivel az ilyen rendszerek finomhangolást követően, más, különböző felhasználásra tervezett MI alapját képezhetik, ezért meghatározásukra 2022-ben a Stanford Egyetem kutatói bevezették a *foundation-model* elnevezést (Bommasani et al., 2022), amelynek szinonimája az AIA által is használt általános célú MI. Általános célú MI képezi a napjainkban már sokak által ismert nagy nyelvi modellek, mint a ChatGPT alapját, de a Google Gemini is ilyen rendszer. A témában lásd még: Goyal (2023); IBM (2021).

[24] Lásd: AIA preambulum (26)–(27) bekezdés.

életviszonyaiban, mint egy hibásan operáló e-mail spam-szűrő. Annak érdekében, hogy ezeket a különbségeket a jogi beavatkozás tükrözze, az AIA különböző kockázati szinteket állít fel, ezekhez pedig eltérő szigorúságú követelményeket és speciális rendelkezéseket állapít meg, melyek igazodnak az MI jelentette kockázatok mértékéhez. A kockázatok mértékét az MI rendszer működésének eredményeként az emberek alapvető jogait, egészségét és a biztonságát,[25] valamint a társadalom egészét potenciálisan fenyegető sérelem súlya, az ebből fakadó hátrányok jelentősége határozza meg. Mindezek alapján négy kockázati kategóriát definiál a Rendelet, melyek egyikébe biztosan be kell sorolni az összes, AIA hatálya alá tartozó MI rendszert.[26] Annak érdekében, hogy e tanulmány fókuszában álló, az AIA célja és rendelkezései közötti aggályos ellentmondás annak teljességében érzékelhető legyen, röviden bemutatom e kategóriákat.

Az elsőt a tiltott MI gyakorlatok köre jelenti, amelyek elfogadhatatlan[27] kockázatot jelentenek a védett értékekre. Ezeket az AIA olyannyira összeegyeztethetetlennek tartja az európai kultúrával és szabadságeszménnyel, hogy semmilyen körülmény között nem engedi használatukat a közösségben. Ide sorolhatók például az MI alapú társadalmi pontozórendszerek.[28] A második kategóriát a magas-kockázatú MI rendszerek alkotják, amelyek értékekre jelentett veszélyeit az AIA egy sor előírással, garanciával és minőségbiztosítási követelménnyel[29] igyekszik mérsékelni. Ezeknek az igen szerteágazó tartalmú rendelkezéseknek való megfelelés komoly kihívásokat fog jelenteni az MI-ben érdekelt gazdasági szereplők számára, hiszen mind technológiai, mind szervezeti megvalósításuk jelentős erőforrásokat igényel. Ráadásul egy részük egyelőre igen homályos, további pontosításokat igényel a jogalkotó részéről.[30] Ennek ellenére az AIA komoly szankciókat[31] helyez kilátásba a szabályok megsértése esetére, ami többek szerint az innováció elfojtásának veszélyével fenyeget (Norgaard, 2024; Business Reporter, 2024; Truby et al., 2022). A nagy-kockázatú MI körben találjuk például a bankok által használt hitelképesség-értékelő MI-rendszereket,[32] vagy a hallgatói csalást észlelő megoldásokat[33] is, de az AIA 6. cikk (1)-(2) bekezdése alapján rendkívül sok MI alapú rendszer nagy-kockázatúnak minősül. A harmadik csoportba a korlátozott kockázatot[34] képviselő megoldások

[25] Lásd: AIA preambulum (1), (7) bekezdés.

[26] Itt fontos megjegyezni, hogy a GDPR-hoz hasonló, igen széles területi és személyi hatályt – lásd az AIA 2. cikkét – kodifikált a jogalkotó. Így még az EU-n kívüli, de az EU-s állampolgárok számára elérhető szolgáltatást nyújtó entitásoknak, pl. MI-rendszereket forgalomba hozóknak, szolgáltatóknak, importőröknek is az AIA rendszeréhez kell igazodniuk, ha szeretnének megjelenni a belső piacon. Az igen kemény extraterritoriális hatálytól az EU véleményem szerint azt a – nem titkolt – eredményt reméli, hogy az ún. Brüsszel-hatásra alapozva ismét nemzetközileg irányadó szabályozást képes kialakítani, ezzel erősítve pozícióját a globális MI versenyben.

[27] Lásd: AIA preambulum (31) bekezdés.

[28] AIA 5. cikk (1) bekezdés c) pont.

[29] Lásd pl. az AIA 8–27. cikkeit.

[30] E kérdések egy részének pontosítására a jogalkotó már a Rendeletben „kötelezettséget vállalt”, azonban a norma magas jogforrási helyzetére és az ésszerű terjedelemre tekintettel nem írta bele az összes szükséges, apró szabályt a kódexbe. Az AIA ugyanakkor számos helyen utalja az EU Bizottság, és annak AIA-t végrehajtó szerve, az AI Office – lásd: AIA 3. cikk (47) bekezdés és 64. cikk – hatáskörébe részletszabályok, egységes gyakorlatok kialakításának feladatát. A Bizottság az egységes közösségi alkalmazás és a normák tartalmának pontos meghatározása érdekében ún. felhatalmazáson alapuló és végrehajtási aktusokat fogadhat el, valamint gyakorlati kódexeket dolgoz ki.

[31] Ld. AIA 99. cikk.

[32] AIA III. melléklet 5. b) pont.

[33] AIA III. melléklet 3. d) pont.

[34] Angolul legtöbbször a „limited-risk” megnevezést használják, ebben a formában nem szerepel a Rendeletben, azonban a jogalkotási folyamat során egységesen rögzült az elnevezés.

tartoznak. Itt találjuk a generatív MI megoldások egy részét, valamint – véleményem szerint kifejezetten helytelenül – a deepfake technológiát is. Az e kategóriára vonatkozó követelményeket a tanulmány szempontjából kiemelt szerepe miatt alább, részletesebben mutatom be. Az utolsó, negyedik kockázati szintet a minimális kockázatú MI alkalmazások – pl. spam szűrők vagy számítógépes játékokba épített MI megoldások (Levine, 2024) – képviselik, amelyek veszélyeit olyannyira elhanyagolhatónak tartja a jogalkotó, hogy a Rendelet nem is határoz meg követelményeket velük kapcsolatban, azonban bátorítja az érintett szolgáltatókat, hogy az ágazati jó gyakorlatokra építő, önkéntes magatartási kódexeket dolgozzanak ki.[35]

4.2. Nyomokban deepfake-et tartalmaz

A deepfake-el kapcsolatos kockázatok régóta ismertek, az utóbbi néhány évben szinte havonta történt a technológiához köthető incidens (Resemble.AI, 2024). Az európai jogalkotás tehát már biztosan e veszélyek tudatában kezdte meg az AIA kodifikációját. Ennek fényében érdekes, hogy a Rendelet végleges szövegében csak elszórva, mindössze négy alkalommal említi a technológiát, amelyből mozaikszerűen áll össze az EU deepfake-doktrínája. Véleményem szerint a jogalkotó világosan látja és rögzíti a deepfake-ben rejlő kockázatok egy részét, ám a kialakított rezsim elégtelen, és nem kezeli megfelelően a technológia által a társadalomra és az egyénekre jelentett valódi veszélyeket, és ezzel jelentősen akadályozza, sőt, veszélybe sodorja az emberközpontú megközelítés érvényesülését. Az alábbiakban kísérletet teszek az AIA-ben található, deepfake-re vonatkozó szabályok összefoglalására és értelmezésre.

Vizsgáljuk meg először az AIA terjedelmes preambulumát, amely több ponton keresztül kristályosítja ki, miért és hogyan szükséges a jogi fellépés az olyan MI megoldásokkal kapcsolatban, ahová a deepfake is tartozik:

> bizonyos MI-rendszerek, amelyek rendeltetése [...] a tartalom létrehozása, különleges kockázatot jelenthetnek a hasonmássággal való visszaélés vagy a megtévesztés szempontjából […] ezért egyedi átláthatósági kötelezettségeket kell alkalmazni. […] Így különösen, a természetes személyeket értesíteni kell arról, hogy MI-rendszerrel állnak interakcióban.[36]

Kitűnik, hogy a legfőbb problémának a személyiségi jogok megsértésének eseteit és a valóság torzítását tartja a jogalkotó. Láthatjuk azt is, hogy a felmerülő problémákat átláthatósági követelmények előírásával kívánja orvosolni, amelyek általános formáját a tájékoztatásában látja. Rögtön ezt követően a Rendelet konkretizálja az említett különleges kockázatok mibenlétét:

> A különböző MI-rendszerek nagy mennyiségű szintetikus tartalmat tudnak előállítani, amelyet illetően egyre nehezebbé válik az emberek számára, hogy megkülönböztessék az ember által előállított és autentikus tartalomtól. E rendszerek széles körű rendelkezésre állása és növekvő képességei jelentős hatással vannak az információs ökoszisztéma integritására és az abba vetett bizalomra, új kockázatokat teremtve a nagyléptékű félretájékoztatás és a manipuláció, a csalás, a személyazonossággal való visszaélés és a fogyasztók megtévesztése tekintetében.[37]

[35] AIA 95. cikk.
[36] AIA preambulum (132) bekezdés.
[37] AIA preambulum (133) bekezdés.

Ugyan e sorok nem szűkítik a tartalomgeneráló MI rendszerek körét csak a deepfake-re, mégis tetten érhetők annak legfőbb sajátosságai és az általam kiemelt példákon keresztül is bemutatott veszélyei. A következő bekezdésben[38] találkozunk először a deepfake terminussal, és választ kapunk arra a kérdésre is, hogy esetében mit takar az átláthatósági követelmény: címkézést és a szintetikus jelleg feltűntetését.

> Az MI-rendszer szolgáltatói által alkalmazott műszaki megoldásokat illetően azon alkalmazóknak, akik MI-rendszert használnak olyan kép-, audio- vagy videotartalom létrehozására vagy manipulálására, amely érzékelhetően hasonlít meglévő személyekre, tárgyakra, helyekre, entitásokra vagy eseményekre, és egy személy számára megtévesztő módon autentikusnak vagy valóságosnak tűnhet (deep fake), egyértelműen és megkülönböztethetően fel kell tüntetniük, hogy a tartalmat mesterségesen hozták létre vagy manipulálták az MI-kimenet megfelelő címkézésével és mesterséges eredetének közzétételével.[39]

A Rendelet itt határozza meg a deepfake definícióját[40] is, amelyet a jogalkotó az AIA-re jellemző módon tágan és általánosan, a változások követéséhez szükséges rugalmasságra törekedve alakított ki. Így a technológia fejlődésének kevésbé kitett fogalommal operál, megelőzve a folyamatos jogszabály-korrekció szükségességét. Az AIA a valóságot tükrözve nem szűkíti csak a mozgóképekre a terminust, és láthatjuk, hogy lényegi elemként tartalmazza a megtévesztő jelleget, ami kulcsfontosságú a szabályozási metodika szempontjából, hiszen pont e tulajdonságát jelöli meg a kockázatok forrásaként.

Az AIA az 50. cikk (4) bekezdésében foglalkozik újra a deepfake-el, meghatározva az ilyen technológia alkalmazóira vonatkozó, a kockázatok mérséklését célzó kötelezettségeket. Alkalmazónak a jogszabály szerint azok a természetes vagy jogi személyek, hatóságok, ügynökségek minősülnek, amelyek a felügyeletük alá tartozó MI-rendszert használják.[41] A használat célját nem tisztázza a Rendelet, azonban egyértelműen kizárja az alkalmazókénti tevékenység lehetséges köréből azokat az eseteket, amikor az MI-rendszert személyes, tehát nem szakmai jellegű célokra használják. A magáncélra való használat tehát nem minősít alkalmazóvá senkit. A bevezető rendelkezésekből kiderül,[42] hogy az alkalmazó általi használat az alkalmazótól eltérő személyeket is érinthet. E szabályokat összeolvasva az a kép áll össze, hogy az alkalmazó olyan végfelhasználó, aki elsősorban saját érdekéből, saját szakmai esetleg üzleti céljaira használja a deepfake technológiát, amelynek működése, eredményei akár másokra is kihatással lehetnek.

Az alkalmazónak mint az átláthatósági kötelezettséget előíró norma címzettjének legfontosabb feladata annak közlése, hogy a tartalmat mesterségesen hozták létre vagy manipulálták.[43] E tájékoztatást az első interakció vagy kitettség alkalmával, egyértelmű és jól megkülönböztethető, akadálymentes módon kell az érintett természetes személyek számára nyújtani.[44] Ameny-

[38] AIA preambulum (134) bekezdés.

[39] Ugyan nem derül ki, hogy a címkézés és a közzététel mit takar, azonban később az 50. cikk (7) bekezdésben megtudjuk, hogy a végrehajtást koordináló MI-hivatalnak segítenie kell uniós szintű gyakorlati kódexek kidolgozását a mesterségesen előállított vagy manipulált tartalom észlelésére és címkézésére vonatkozó kötelezettségek hatékony végrehajtásának elősegítése érdekében. Bízzunk benne, hogy minél előbb eleget tesznek e kötelezettségüknek.

[40] Ezt ismétli meg a 3. cikk 60. pontjában.

[41] AIA 3. cikk 4. pont.

[42] AIA preambulum (13) bekezdés.

[43] AIA 50. cikk (4) bekezdés.

[44] AIA 50. cikk (5) bekezdés.

nyiben azonban a deepfake tartalom „nyilvánvalóan művészeti, kreatív, szatirikus, fiktív vagy hasonló mű vagy program részét képezi”, azaz például egy mozifilmben látható, úgy elegendő a megjelenést és az élvezeti értéket nem akadályozó közlés. Azt, hogy e tájékoztatást a gyakorlatban pontosan hogyan, milyen formában kell nyújtani, nem tudjuk meg az AIA-ből, mivel kódex a Bizottságra hagyja ennek kidolgozását.[45]

Az AIA rendelkezéseinek betartását, így az átláthatósági követelmények teljesítését is többszintű hatósági kontroll biztosítja. A Rendelet minden tagállamban elrendeli a jogszabály végrehajtásáért dedikáltan felelős, független és pártatlan piacfelügyeleti hatóság létrehozását vagy kijelölését[46] és alapvetően ezekre a szervekre bízza az AIA-nek való megfelelés ellenőrzését.[47] Az egységes közösségi felügyelet és biztonság megteremtése érdekében a Rendelet kiterjeszti[48] az EU piacfelügyeleti irányelvének[49] hatályát is az AIA hatálya alá tartozó MI rendszerekre és gazdasági szereplőkre, így az alkalmazókra is. Amennyiben az illetékes nemzeti hatóságnak elegendő oka van úgy ítélni, hogy egy MI-rendszer a személyek egészségére, biztonságára, illetve alapvető jogaira nézve kockázatot jelent, kötelező elvégeznie annak értékelését, és vizsgálni annak megfelelését az AIA-ben meghatározott követelményeknek.[50] Annak érdekében, hogy egy incidens se maradjon rejtve, az AIA bárkinek biztosítja, hogy panaszokat tegyen a releváns piacfelügyeleti hatóságnál, ha úgy véli, megsértették a kódex rendelkezéseit.[51] Ha az alapjogokat is veszélyeztető kockázatot azonosítanak – ami a deepfake esetében nem ritka –, úgy a piacfelügyeleti hatóságnak értesíteni kell az alapvető jogok védelmét felügyelő nemzeti hatóságokat is,[52] és velük együttműködve szükséges elvégezniük az értékelést. Elképzelhető, hogy a meg nem felelés nem csak egy ország területére korlátozódik, ilyenkor a nemzeti hatóságnak késedelem nélkül tájékoztatnia kell a Bizottságot és a többi államot is.[53]

Abban az esetben, ha a hatóság úgy találja, hogy az MI-rendszer nem felel meg a Rendeletben megállapított szabályoknak, haladéktalanul elő kell írnia a gazdasági szereplőnek,[54] hogy rövid határidőn belül tegyen meg minden szükséges korrekciós intézkedést a megfelelés érdekében, vagy vonja ki az MI-rendszert a forgalomból, esetleg hívja vissza azt. Ha az érintett ezt követően nem tesz eleget a korrekciós intézkedéseknek, úgy a hatóságnak meg kell hoznia minden megfelelő átmeneti intézkedést az MI-rendszer nemzeti piacon való jelenlétének megtiltása, korlátozása, illetve a MI-rendszer forgalomból való kivonása vagy visszahívása érdekében.[55] Erről az intézkedéséről és az ügy minden további érdemi részletéről értesíteni kell a

[45] AIA 96. cikk (1) bekezdés d) pont.

[46] AIA 70. cikk (1) bekezdés.

[47] Egyes MI-rendszerek felügyeletét az általános szabálytól eltérően a Rendelet specializált piacfelügyeleti hatóságokra bízza, lásd pl. az AIA 74. cikk (3), (6), (8) és (9). bekezdéseit. Az ún. általános célú MI-rendszerek – ide tartozik a korábban említett, a deepfake előállításban új távlatokat nyitó generatív MI megoldások jelentős része is – esetében a felügyeletet az MI-hivatalra bízza, amennyiben az általános célú MI-modell és a rendszer kifejlesztését ugyanazon szolgáltató végezte. Lásd az AIA 75. cikk (1) bekezdését.

[48] AIA 74. és 75 cikkei.

[49] Az Európai Parlament és a Tanács (EU) 2019/1020 rendelete (2019. június 20.) a piacfelügyeletről és a termékek megfelelőségéről, valamint a 2004/42/EK irányelv, továbbá a 765/2008/EK és a 305/2011/EU rendelet módosításáról.

[50] AIA 79. cikk (2) bekezdés.

[51] AIA 85. cikk.

[52] Lásd: AIA 79. cikk (2) bekezdés, 77. cikk (1) bekezdés.

[53] AIA 79. cikk (3) bekezdés.

[54] AIA 79. cikk (2) bekezdés.

[55] AIA 79. cikk (5) bekezdés.

Bizottságot és a tagállamokat. Ha az eljárásra az 50. cikknek való meg nem felelés miatt került sor, azt külön is jelezni kell.[56] Utóbbi rendelkezés megerősíti, hogy az AIA kiemelt figyelmet fordít a szintetikus tartalmakra, így a deepfake-re vonatkozó szabályok betartására, és minden végrehajtó szervet igyekszik naprakészen tartani a visszaélésekkel kapcsolatban.

4.3. A hatályos deepfake rezsim hiányosságai és aggályai

Véleményem szerint az AIA több átgondolatlan döntést tartalmaz, deepfake-re fókuszáló rendelkezései jelen formájukban nem szolgálják maradéktalanul – sőt inkább veszélyeztetik – az emberi jogok védelmét és az etikai alapelvek érvényre juttatását, ezzel a Rendelet origóját, az emberközpontúságot. A problémák gyökere a szabályozási metodika helytelen kiindulási pontjából fakad, az AIA ugyanis rossz irányból tekint a technológiára: a rezsimet a deepfake jogszerű, hasznos, „kellemes" felhasználása által jelentett kockázataihoz igazítja, ami azonban a teljes „deepfake-kibocsátásnak" csak elenyésző része. Ha a jogalkotó tekintettel lett volna a jóval gyakoribb visszaélésszerű és káros alkalmazásokra, és ezeket veszi alapul a kockázatok felmérésekor, akkor szigorúbb, pontosabban kidolgozott, egyben emberközpontúbb jogszabályt alkotott volna. Természetesen érthető az a megfontolás is, miszerint a közösségi jogalkotás a tagállami szuverenitást tiszteletben tartva a nemzeti büntetőjogokra bízza az ártó felhasználás kezelését – hiszen ezek az esetek kivétel nélkül valamilyen bűncselekményt valósítanak meg. A veszélyek súlyára tekintettel azonban minden lehetséges eszközt meg kellett volna ragadnia azok megelőzése érdekében, így az ilyen deepfake létrehozására alkalmas MI rendszerek vagy tartalmak tiltásáig is el lehetett volna menni.

A deepfake olyan károkat tud okozni, amelyek messze meghaladják azt a kockázati szintet, ahogyan az AIA jelenleg kezeli, éppen ezért úgy vélem, elhibázott a „korlátozott kockázatú" kategóriába való besorolása is. Az AIA által az ide tartozó MI rendszerekre meghatározott rendelkezésekkel elérni kívánt jogvédelmi szint nem tükrözi a technológia által jelentett veszélyek súlyát, sőt, messze alulmúlja azt. Ha az EU a veszélyeket helyén kezelné, bizonyos deepfake alkalmazási területeket, felhasználási eseteket és alkalmazói magatartásokat akár az elfogadhatatlan kockázatot jelentő kategóriába is sorolhatott volna, hiszen a szelektív tiltásra számos nemzetközi példát ismerünk.[57] Ehelyett logikailag (is) felborítva a kockázatalapú megközelítést, alábecsült egy, a Rendelet által védett értékekre és a társadalomra potenciális és eltúlozhatatlan fenyegetést jelentő technológiát.

További – egyelőre megválaszolatlan – kérdéseket vet fel, hogy a deepfake generáló MI-rendszerek alkalmazóinak miért csak a szakmai jellegű tevékenységük során kell az előírt átláthatósági követelményeknek eleget tenni? Egyáltalán hol a határ a szakmai és a nem szakmai felhasználás között? Miért sugallja a Rendelet, hogy az „otthoni" felhasználás veszélytelenebb, amikor éppen ennek eredménye a rengeteg, kifejezetten jogsértő deepfake? Reális, hogy valaki üzleti tevékenységének vagy szakmájának részeként készít jogsértő tartalmakat, melyeken a Rendeletnek megfelelve egyértelmű módon jelzi azok valótlanságát? Aligha. Épp ellenkezőleg: aki ilyesmire vetemedik, mindent bevet annak érdekében, hogy névtelenségét megőrizze, és soha ne tudják összefüggésbe hozni az általa létrehozott deepfake-el. Az pedig, hogy ilyen személyektől, csoportoktól a jogszabályok betartását várjuk, kissé paradox gondolkodásra vallana.

[56] AIA 79. cikk (6) bekezdés d) pont.

[57] Lásd az 5. fejezetet.

4.4. Biztató jelek

A sok nyitott kérdés súlyos, mielőbb kezelendő problémákat takar. Az AIA-ben azonban mégis találunk olyan, némiképp megnyugtató rendelkezéseket, amivel a jogszabály igyekszik felvenni a harcot a veszélyekkel szemben. A szintetikus hangot, képet, videót vagy szöveget[58] létrehozni képes rendszerek szolgáltatóinak ugyanis kötelező e kimenetet géppel olvasható formátumban megjelölni úgy, hogy azok mesterségesen létrehozottként vagy manipuláltként észlelhetők legyenek.[59] E jelölésnek a lehető legmodernebb megoldásokon[60] kell alapulnia és interoperábilisnak kell lennie, aminek célja, hogy könnyen azonosítani lehessen a szintetikus tartalmat. Ugyan az AIA nem definiálja a szintetikus tartalom fogalmát és körét, azonban abból kiindulva, hogy az átláthatósági követelmény – adott esetben a számítógépes megjelölés – kiterjed mesterségesen manipulált tartalmakra is, véleményem szerint az is következik, hogy e kötelezettség teljesítése érinti a teljes deepfake palettát, akár modern – generatív MI –, akár primitívebb technolóiával – arccserélés – hozták létre azokat. A Rendelet szerint[61] szolgáltatónak minősül minden olyan természetes vagy jogi személy, hatóság, ügynökség vagy egyéb szerv, amely MI-rendszert vagy általános célú MI-modellt fejleszt vagy fejleszttet, és a saját neve vagy védjegye alatt – akár fizetés ellenében, akár ingyenesen – forgalomba hozza, üzembe helyezi. Összeolvasva tehát: az AIA hatálya alatt elvileg nem keletkezhet jogszerűen olyan deepfake, amiről legalább számítógéppel ne volna megállapítható „valódi kiléte”, azonban azt, hogy az ilyen tartalomról mindenki számára felfogható módon, egyértelműen jelezzék valótlanságát, a Rendelet nem teszi kötelezővé – legalábbis a személyes célú, nem szakmai alkalmazás esetében. Tekintettel arra, hogy az ártó célú deepfake készítés és felhasználás az esetek jelentős részében éppen a magánhasználat következménye, a rendelkezések kockázat-mérséklő ereje és hatása erősen megkérdőjelezhető.

Érdemes kiemelni a jogalkotó két másik, igen átgondolt döntését is, amelyek arról tanúskodnak, hogy az említett hibák mellett is felismerték a deepfake jelentette veszélyeket. Az első a már taglalt átláthatósági követelmények kiterjesztése a szabad és nyílt forráskódú licencek által kibocsátott MI-rendszerekre. Az AIA alapvetően nem alkalmazandó a mindenki számára hozzáférhető megoldásokra, azonban a 2. cikk (12) bekezdése értelmében a hatálya mégis kiterjed az ilyen MI-rendszerekre, amennyiben azokat az 50. cikk hatálya alá tartozó MI-rendszerként[62] hozzák forgalomba, vagy helyezik üzembe. Tekintettel arra, hogy számos deepfake létrehozására képes alkalmazás nyílt forráskóddal rendelkezik, így e rendelkezés kvázi a védvonalak kiterjesztéseként is értelmezhető, ami megnövelheti az AIA hatékonyságát az ártó képességgel is rendelkező MI alkalmazásokkal szemben.

Végső esetben pedig az AIA 99. cikke komoly összegű bírságokat helyez kilátásba a Rendeletnek való meg nem felelés, így a szolgáltatókra és az alkalmazókra vonatkozó, az 50. cikk szerinti átláthatósági kötelezettségek be nem tartása esetére is. Tekintettel arra, hogy ennek összege akár 15.000.000 euró, vagy vállalkozás esetén az előző pénzügyi év teljes globális éves árbevételének legfeljebb 3%-át kitevő összeg is lehet, igen komoly elrettentő erővel bír.

---

[58] Ahogyan már utaltam rá, ilyen megoldásokkal hozzák létre napjainkban a deepfake-ek túlnyomó többségét.

[59] AIA 50. cikk (2) bekezdés.

[60] Az AIA preambulum (133) bekezdés példálózva sorol föl néhány műszaki megoldást, így „például vízjeleket, metaadat-azonosításokat, a tartalom eredetének és hitelességének bizonyítására szolgáló kriptográfiai módszereket, naplózási módszereket, ujjlenyomatokat vagy egyéb technikákat”.

[61] AIA 3. cikk 3. pont.

[62] Az 50. cikk (4) bekezdése szerint a deepfake-ek is ide tartoznak.

A bírságok kiszabásának módjával kapcsolatban az AIA tiszteletben tartja a tagállamok eltérő közigazgatási jogrendszereit, így csak a kereteket szabja meg:[63] elsősorban a közigazgatási bírságokra vonatkozó szabályok alkalmazását írja elő azzal, hogy a bírságokat – megfelelő eljárási garanciák mellett – az illetékes nemzeti bíróságoknak, vagy adott esetben más, erre feljogosított szerveknek kell kiszabnia. A kódex ugyanakkor közös rendező elvként köti ki, hogy a szankciók alkalmazásának minden tagállamban azonos hatással kell járnia, és biztosítani kell az azokkal szembeni hatékony jogorvoslatokat. További közösségi kontrollt jelent, hogy a tagállamoknak a végrehajtási intézkedésekre vonatkozó belső szabályokról haladéktalanul, de legkésőbb az alkalmazás kezdőnapjáig értesíteni kell a Bizottságot. Mindezek mellett meggyőződésem, hogy a deepfake-et etikus célokra készítők mindent meg fognak tenni a bírság elkerülése érdekében, azonban ahogy említettem, nem ők jelentik a fő problémát.

5. Kitekintés

Bár az EU nyíltan törekszik a globális MI szabályozás irányainak meghatározására, egyelőre az USA és Kína mögött kullog. E két hatalom különböző okokból, de igencsak elhúzott az MI versenyben, ami egyben azt is jelenti, hogy érdemes figyelemmel követni szabályozási tevékenységüket, ezért a tanulmány célját szem előtt tartva kiemelem az USA deepfake megközelítésének fontosabb mozzanatait, illetve néhány nemzetközi példával is árnyalom az EU szabályozási politikáját a területen. A tavalyi év jogalkotási eseményei bizakodásra adtak okot, mivel az Egyesült Államok elnöke által deklarált MI megközelítés részeként született néhány, a szintetikus tartalom – így a deepfake-ek – helyzetét is érintő norma. Az USA és az EU végig fej-fej mellett haladtak a saját átfogó MI szabályozási kereteik kialakításával, de végül az AIA-t röviddel megelőzve született[64] az az elnöki rendelet,[65] ami szövetségi szinten kodifikálta a technológiával kapcsolatos egyes kérdések kereteit.

A rendelet már a céljai között rögzíti, hogy: „az MI felelőtlen használata súlyosbíthatja az olyan társadalmi károkat, mint a dezinformáció".[66] A jogszabály ugyan nem operál fogalmi szinten deepfake-el, azonban több olyan rendelkezést tesz, amelyek implicit módon, de kiterjednek az ilyen MI által generált tartalmakra, és azok veszélyeinek kezelésére is. Az elnök utasította kereskedelmi minisztert, hogy a kockázatok csökkentése érdekében építse ki mesterséges intelligencia-rendszerek által előállított szintetikus tartalom azonosítására és címkézésére alkalmas képességeket. A rendelet értelmében több kormányzati hivatalnak együttműködve kell a meglévő gyakorlatokat összegző, tudományosan alátámasztott, további standardok kialakítását megalapozó jelentést készíteni[67] olyan témakörökben, mint a szintetikus tartalmak felismerése, azok megjelölése – például digitális vízjellel –, valamint a generatív AI által, valódi személyeket ábrázoló non-konszenzuális szexuális tartalmak előállításának megelőzése.[68] Láthatjuk, hogy az elnöki rendelet hatóköre számos hasonlóságot mutat az AIA-vel, szinte azonos fókuszpon-

[63] Lásd az AIAI 99. cikk (1)-(2) és (9)-(10) bekezdés rendelkezéseit.

[64] A két jogszabály összehasonlításáról ld.: Boone (2024).

[65] Executive Order 14110 of Oct 30, 2023 - Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence [E.O].

[66] E.O. 14110. Sec. 1.

[67] E.O. 14110. Sec. 4.5. a).

[68] E.O. 14110. Sec. 4.5. a) (ii)-(iv).

tokat határozva meg a szintetikus média kezelésével kapcsolatban. Az említett jelentést[69] igen széles társadalmi és piaci egyeztetésre alapozták, a visszajelzések[70] számos esetben kiemelték a deepfake általam is tárgyalt veszélyeit, aminek fényében bizakodhatunk, hogy a jövőbeli esetleges szövetségi szintű szabályozás megfelelően fogja kezelni a technológia által jelentett kockázatokat. Megemlítendő, hogy az elmúlt években több, kifejezetten deepfake-specifikus törvényjavaslatot (Graham, 2024; Clarke, 2023; Morelle, 2023) is beterjesztettek a Kongresszus elé, ám – legalábbis e tanulmány lezárásáig – nem került sor azok elfogadására.

Habár a szövetségi kodifikáció még várat magára, néhány állam már komoly lépéseket tett a deepfake-ek káros felhasználásának visszaszorítására (Graham, 2024): Texasban bűncselekménnyé nyilvánították az olyan megtévesztő deepfake videók készítését, amelyeknek célja a választás eredményének befolyásolása, míg Új Mexikóban vagy Indianában kötelezővé tették a politikai kampányban használt deepfake tartalmak megjelölését. Az államok eltérően közelítik meg a szexuális tartalmú deepfake-ek helyzetét. A bosszúpornót – így annak mesterségesen előállított változatát is – 46 államban büntetik (Hao, 2021), de Lousianaban vagy Florida államban például csak a kiskorúakat ábrázoló deepfake tartalmakat tették büntethetővé (Graham, 2024), míg négy állam általános jelleggel tiltja az MI által létrehozott, non-konszenzuális pornográfiát (Donegan, 2023). Világszerte ugyancsak vegyes a helyzet. Japánban például éppen a közelmúltban tiltották be az olyan szexuális tartalom előállítását, amelyeket a szereplők beleegyezése nélkül készítenek (MTI, 2023). Az Egyesült Királyságban ennél is tovább mennek: az idei évtől az explicit deepfake-ek létrehozói korlátlan összegű bírsággal néznek szembe, de akár börtönbe is kerülhetnek, ha az eredmény nagyobb nyilvánosság előtt megosztásra kerül (Cooney, 2024).

Már szóba került, hogy az ártó deepfake tartalmakat rendszerint nem az internet sötét sarkaiban eldugva találjuk, hanem egyszerű Google kereséssel hozzáférhetők. Éppen ezért magam is osztom azok véleményét, akik úgy tartják (Lawelle, 2024), hogy a deepfake-ek visszaszorításában nemcsak az államoknak, de a technológiát fejlesztő, valamint az annak készítéséhez és terjesztéséhez szükséges ökoszisztémát biztosító – így a megosztási felületeket, online tereket kontrolláló – piaci szereplőknek is kiemelt szerepe van. Vitán felül áll, hogy e felelősség elsősorban a legnagyobb közösségi médiát és egyéb digitális szolgáltatást nyújtó tech-óriásokat terheli, akiket az EU digitális piacait szabályozó rendelete „kapuőröknek" nevez.[71] Ilyen cég például a Meta, az Amazon vagy a Google tulajdonos Alphabet (Schmidt, 2024). Rajtuk kívül kiemelt szerepe van az online kereskedelmet facilitáló vállalkozásoknak is. Fontosságukat jól mutatja az NBC News tavalyi esettanulmánya (Tenbarge, 2023), mely szerint az egyik legnagyobb, szexuális tartalmú deepfake-et is megosztó oldalon az akár személyre szabott tartalmakat a két vezető fizetési szolgáltató, a Visa és a Mastercard rendszerén keresztül is megvásárolhatják az ügyfelek. Mindezt annak ellenére, hogy a Visa üzletszabályzata például kifejezetten tiltja, hogy a cég termékeivel olyan tartalomért fizessenek, ami beleegyezés nélküli szexualitást ábrázol.

[69] A dokumentum e tanulmány lezárásának idején még nem készült el, tervezete itt olvasható: https://shorturl.at/ygFrT

[70] A visszajelzések itt olvashatóak: https://shorturl.at/bivHU

[71] Lásd az Európai Parlament és a Tanács (EU) 2022/1925 Rendelete (2022. szeptember 14.) a digitális ágazat vonatkozásában a versengő és tisztességes piacokról, valamint az (EU) 2019/1937 és az (EU) 2020/1828 irányelv módosításáról (digitális piacokról szóló jogszabály) [DMA] 3. cikkét.

6. Záró gondolatok

Bízom benne, hogy tanulmányom végére érve egyértelművé váltak a deepfake megjelenésével az emberiségre zúduló óriási kihívások és veszélyek, amelyeknek bárki, legyen akármilyen felkészült, áldozatául eshet. Az AIA rendelkezéseit tanulmányozva esetleg az a benyomásunk alakulhatott ki, mintha a jogalkotó némiképp felelőtlenül szemet hunyt volna a veszélyek felett. Bár a kódex törekedett arra, hogy ne hagyja figyelmen kívül a technológiát, rendelkezései – melyek a maguk nemében helytállók – mégis olyan oldalról ragadják meg a jelenséget, ahol az emberi jogok és a társadalom biztonsága a talán még komoly jogi beavatkozás nélkül is védelmet élvezne. Úgy gondolom, ha valóban szeretné érvényesíteni az emberközpontúságot, a Rendeletnek észlelnie kellene a deepfake visszaélésszerű használatával felmerülő kockázatok teljes körét, és keményen reflektálnia kellene azokra. A fellépés a deepfake-re vonatkozó szabályok szigorításával, magasabb kockázati kategóriába való besorolásával, speciális biztonsági intézkedések bevezetésével, végső soron pedig – a nem példátlan – kategorikus tiltással is megvalósítható volna.

Fontos leszögeznünk azonban, hogy a deepfake jelentette kockázatokat jogi előírásokkal csak mérsékelni lehet, kizárni még az AIA-nél jóval szigorúbb rendelkezésekkel sem. Lássuk be, a jog jelen tudásunk szerint eszköztelen egy, a bűnüldöző hatóságok számára is alig ismert, az interneten könnyen és kvázi korlátok nélkül terjedő technológiával szemben. Önmagában a rossz célú használat tiltása, az AIA-ben kodifikált átláthatósági követelményekhez hasonló szabályok előírása nem fogja teljesen megoldani az aggasztó helyzetet. Aki ártó szándékkal hoz létre ilyen tartalmakat, az nemhogy nem fogja feltüntetni azok deepfake-jellegét, hanem minden tőle telhetőt megtesz annak érdekében, hogy eltűnjön a föld színéről.

A bemutatott veszélyek önmagukban is jelentős kockázatokat jelentenek az emberekre, alapjogaikra és a társadalom egészére nézve, azonban a helyzetet súlyosbítja a deepfake-el és a mesterséges intelligenciával kapcsolatban fennálló, általános társadalmi tudatlanság és homály. Nem csoda, hogy „a deepfake által generált fake news-t nem is a dezinformációs szándékkal terjesztő felhasználók terjesztik a legnagyobb mértékben, hanem azok, akiket megtévesztett a manipulált tartalom.” (Gosztonyi & Lendvai, 2024, 44). Ha eljutunk oda, hogy valaki egy ilyen videót látva megkérdőjelezi annak hitelességét, akkor már egy lépéssel közelebb kerültünk a technológia uralásához, és nem vesztettük el a harcot. Éppen ezért véleményem szerint a káros deepfake tartalmak veszélyeinek csökkentéséhez sokkal nagyobb szükség volna átgondolt és szisztematikus, az MI-tudatosságot növelő programokra és stratégiára, mintsem nehezen végrehajtható kikényszeríthető jogszabályokra. Meggyőződésem, hogy az ártó deepfake-ek készítése és terjesztése ellen a jogalkotás, a piac és a civil szféra közös, összehangolt fellépése szükséges, amit átfogó társadalmi tudásbővítés egészít ki. Így talán van esélyünk arra, hogy a volánnál maradjunk az MI korában.

Hivatkozások

Ahmed, Z. (2024. január 22.). *Fake Joe Biden robocall: AI-Generated Election Interference.* Resemble AI. Online: https://tinyurl.com/2juukrxx

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact.* Deeptrace.

Bitport (2024. április 26.). *Ezúttal egy iskolaigazgatót hurcoltak meg deepfake miatt.* Online: https://tinyurl.com/262n754a

Black, J., & Fullerton, C. (2020). Digital Deceit: Fake News, Artificial Intelligence, and Censorship in Educational Research. *Open Journal of Social Sciences*, 8(7), 71–88. https://doi.org/10.4236/jss.2020.87007

Bodnár Z. (2020. február 21.). *Először vetettek be választási kampányban deepfake videókat*. Qubit. Online: https://tinyurl.com/5av66hza

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2022). *On the Opportunities and Risks of Foundation Models* (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

Boone, T. S. (2024. szeptember 9.). *An Examination Of Certain Key Features Of The New White House Executive Order On Artificial Intelligence*. Corvinus University of Budapest. Online: https://unipub.lib.uni-corvinus.hu/10301/

Business Reporter (2024. május 29.). *The EU AI Act: Necessary regulation or a barrier to innovation?* Online: https://tinyurl.com/24rhmtvt

Carpenter, P. (2024. november 6.). *Deepfakes didn't disrupt the election, but they're changing our relationship with reality*. The Hill. Online: https://tinyurl.com/58d7cuj4

Chadha, A., Kumar, V., Kashyap, S., & Gupta, M. (2021). Deepfake: An Overview. In P. K. Singh, S. T. Wierzchoń, S. Tanwar, M. Ganzha, & J. J. P. C. Rodrigues (Szerk.), *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security* (pp. 557–566). Springer Singapore. https://doi.org/10.1007/978-981-16-0733-2_39

Chen, H., & Magramo, K. (2024. február 4.). *Finance worker pays out $25 million after video call with deepfake 'chief financial officer'*. CNN. Online: https://tinyurl.com/ms8xbmyx

Cheng, J. (2013. július 12.). *Bruce Lee Controversially Resurrected for Johnnie Walker Ad*. Time. Online: https://newsfeed.time.com/2013/07/12/bruce-lee-resurrected-for-whisky-ad/

Clarke, Y. D. (2023. szeptember 21.). *Clarke Leads Legislation to Regulate Deepfakes*. Congresswoman Yvette Clarke. Online: https://bit.ly/3ONDKbE

Cooney, C. (2024. április 16.). *Creating sexually explicit deepfakes to become a criminal offence.* BBC. Online: https://www.bbc.com/news/uk-68823042

Council of the EU (2024. május 21.). *Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI*. Online: https://tinyurl.com/5d7w8nhc

Damiani, J. (2019. szeptember 3.). *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*. Forbes. Online: https://shorturl.at/Xzs62

Davies, S. (2024. január 12.). *Generative AI and deepfakes. How AI will create disinformation.* Online: https://thumos.uk/generative-ai-and-deepfakes/

De Luce, D. (2024. október 22.). *Can you spot the celebrity „deepfakes" in a new ad warning against election disinfo?* NBC News. Online: https://tinyurl.com/444x497f

Devlin, K., & Cheetham, J. (2023. március 24.). *Fake Trump arrest photos: How to spot an AI-generated image*. BBC News. Online: https://www.bbc.com/news/world-us-canada-65069316

Donegan, M. (2023. március 13.). *Demand for deepfake pornography is exploding. We aren't ready for this assault on consent*. The Guardian. Online: https://tinyurl.com/mrk95byt

Dredge, S. (2016. március 17.). *Five of the best face swap apps*. The Guardian. Online: https://tinyurl.com/ywvwxbau

Elliott, V. (2024. május 30.). *The WIRED AI Elections Project*. Wired. Online: https://bit.ly/3OLH3jy

Europol (2024). *Facing reality? Law enforcement and the challenge of deepfakes: An observatory report from the Europol innovation lab*. Publications Office. Online: https://doi.org/10.2813/158794

Európai Bizottság (2019). *Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan*. Publications Office. Online: https://doi.org/10.2759/428483

Ewe, K. (2023. december 28.). *Elections Around the World in 2024*. Time. Online: https://bit.ly/41qIwmP

Feiner, L. (2024. augusztus 21.). *Telecom will pay $1 million over deepfake Joe Biden robocall*. The Verge. Online: https://tinyurl.com/mu9w9m6e

Gosztonyi G., & Lendvai G. (2024). Deepfake és dezinformáció. Mit tehet a jog a mélyhamisítással készített álhírek ellen? *Médiakutató*, 25(1), 41–49. https://doi.org/10.55395/MK.2024.1.3

Goyal, M. (2023. március 8.). *What is generative AI, what are foundation models, and why do they matter?* IBM Blog. Online: https://tinyurl.com/zzae95xu

Graham, M. M. (2024. június 26.). *Deepfakes: Federal and state regulation aims to curb a growing threat*. Thomson Reuters Institute. Online: https://tinyurl.com/235zahsf

Guterman, D. (2024. május 29.). *AI and the 2024 Elections*. Harvard Kennedy School. Ash Center for Democratic Governance and Innovation. Online: https://bit.ly/4gnwTkO

Hacker, P. (2023). *What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models*. Verfassungsblog. Online: https://bit.ly/49vjeFX

Hao, K. (2020a. október 9.). *Inside the strange new world of being a deepfake actor*. MIT Technology Review. Online: https://tinyurl.com/fz9dswrb

Hao, K. (2020b. október 20.). *A deepfake bot is being used to "undress" underage girls*. MIT Technology Review. Online: https://tinyurl.com/22ej4fcd

Hao, K. (2021. február 12.). *Deepfake porn is ruining women's lives. Now the law may finally ban it*. MIT Technology Review. Online: https://tinyurl.com/yc26rvht

Herke Cs. (2023). Deepfake: Áldás vagy átok? *Pro Futuro*, 13(1), 157–178. https://doi.org/10.26521/profuturo/2023/1/13334

Holdsworth, J., & Scapicchio, M. (2024. június 17.). *What is Deep Learning?* IBM. Online: https://www.ibm.com/topics/deep-learning

Holliday, C. (2021). Rewriting the stars: Surface tensions and gender troubles in the online media production of digital deepfakes. *Convergence: The International Journal of Research into New Media Technologies*, 27(4), 899–918. https://doi.org/10.1177/13548565211029412

Hood, C. (2021. augusztus 29.). *How Deepfake Technology Can Change The Movie Industry*. ScreenRant. Online: https://tinyurl.com/4mfb9mb2

Huang, K. (2023. április 8.). *Why Pope Francis Is the Star of A.I.-Generated Photos*. The New York Times. Online: https://tinyurl.com/ycx2ss2a

Hurst, L. (2023. október 20.). *How AI is driving an explosive rise in deepfake pornography*. Euronews. Online: https://tinyurl.com/4ssm7cy6

IBM (2021. február 9.). *What are foundation models?* IBM Research. Online: https://bit.ly/4i-CV150

Jain, J. (2023. április 29.). *The AI Intimacy Trap – How Persuasion Machines Can Lead To Societal Collapse And What To Do About It*. Hotel AI, Marketing, Tech and Loyalty. Online: https://tinyurl.com/5f4y9rm3

Kan, M. (2019. május 17.). *This AI Can Recreate Podcast Host Joe Rogan's Voice To Say Anything*. PCMAG. Online: https://tinyurl.com/m63cxc5d

Kelleher, K. (2023. augusztus 10.). *Revenge Porn and Deep Fake Technology: The Latest Iteration of Online Abuse*. Boston University. Online: https://tinyurl.com/amvk85x6

Koltay A. (2020). A hazugságok alkotmányos védelme? *Jogtudományi Közlöny*, 75(7–8), 302–313. Online: https://real-j.mtak.hu/13974/21/JK_2020_7_8.pdf

Lawelle, B. (2024. február 2.). *We Are In The Middle Of An AI Deepfake Porn Crisis*. Junkee. Online: https://junkee.com/ai-deepfake-porn-crisis-explained/356764

Levine, A. (2024. június 19.). *The EU AI Act: Key provisions and future impacts*. Thoropass. Online: https://thoropass.com/blog/compliance/eu-ai-act/

Linn, G. (2024. január 26.). *Mia Janin took own life after bullying—Inquest*. BBC. Online: https://www.bbc.com/news/articles/cn0nd1gnj4lo

Maddocks, S. (2020). 'A Deepfake Porn Plot Intended to Silence Me': Exploring continuities between pornographic and 'political' deep fakes. *Porn Studies*, 7(4), 415–423. https://doi.org/10.1080/23268743.2020.1757499

Mahdawi, A. (2023. április 1.). *Nonconsensual deepfake porn is an emergency that is ruining lives*. The Guardian. Online: https://tinyurl.com/ycyvz44z

Molnár G. (2024. augusztus 19.). *Donald Trump elfogadta Taylor Swift „támogatását" – AI-képeken a rajongók*. Index. Online: https://tinyurl.com/2mdnyet3

Morelle, J. (2023. május 5.). *Congressman Joe Morelle Authors Legislation to Make AI-Generated Deepfakes Illegal*. Online: https://tinyurl.com/hzz2fk8n

MTI (2023. május 2.). *Japánban betiltják a beleegyezés nélkül készített szexuális tartalmú felvételeket*. Maszol. Online: https://tinyurl.com/2ectrs5n

Naezer, M., & van Oosterhout, L. (2021). Only sluts love sexting: Youth, sexual norms and non-consensual sharing of digital sexual images. *Journal of Gender Studies*, 30(1), 79–90. https://doi.org/10.1080/09589236.2020.1799767

Narvali, A. M., Skorburg, J. A. (Gus), & Goldenberg, M. J. (2023. november 28.). *Cyberbullying girls with pornographic deepfakes is a form of misogyny*. The Conversation. Online: https://tinyurl.com/y2389rhx

Németh Á. (2024. március 24.). *Giorgia Meloni* kártérítést kér a róla terjesztett *deepfake videók miatt*. Index. Online: https://tinyurl.com/c3wvn3my

Norgaard, J. G. (2024. augusztus 27.). *Council Post: The US Innovates, The EU Regulates: How Can The EU Change This Narrative?* Forbes. Online: https://tinyurl.com/yckrytrm

Payne, L. (2024. augusztus 11.). *Deepfake*. Britannica. Online: https://bit.ly/41IJKdt

Resemble.AI. (2024, szeptember 26). *AI Safety: Deepfake Incident Report*. Online: https://bit.ly/3ZKkCSb

RTL (2024. május 16.). *Gyermekpornográfiának is minősülhet az általános iskolai osztálytársról készült deepfake videó* [Video]. Online: https://tinyurl.com/4ff5h9cu

Ryan-Mosley, T. (2023. december 1.). *A high school's deepfake porn scandal is pushing US lawmakers into action*. MIT Technology Review. Online: https://tinyurl.com/42bu9ttn

Schmidt, J. P. (2024. május 31.). *Digital Markets Act: European Commission designates further gatekeeper and core platform services*. NOERR. Online: https://tinyurl.com/35sr5m4x

Security Hero (2023). *2023 State Of Deepfakes: Realities, Threats, And Impact*. Online: https://bit.ly/4fb3ERk

Sensity (2024). *The State of Deepfakes 2024*. Online: https://sensity.ai/reports/

Shepardson, D. (2024a. május 23.). *US political consultant indicted over AI-generated Biden robocalls*. Reuters. Online: https://tinyurl.com/hkmx7m5p

Shepardson, D. (2024b. szeptember 26.). *Consultant fined $6 million for using AI to fake Biden's voice in robocalls*. Reuters. Online: https://tinyurl.com/3t7rbwss
Somers, M. (2020. július 21.). *Deepfakes, explained*. MIT Sloan. Online: https://bit.ly/3Dhvj5B
Speechify. (2024). *Real-Time AI Dubbing with Voice Preservation*. Online: https://bit.ly/3B9c3XD
Spring, M. (2024. március 4.). *Trump supporters target black voters with faked AI images*. BBC. Online: https://www.bbc.com/news/world-us-canada-68440150
Taylor, M. (2024. október 3.). *An AI Deepfake Could Be This Election's November Surprise*. Time. Online: https://time.com/7033256/ai-deepfakes-us-election-essay/
Tenbarge, K. (2023. március 27.). *Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy*. NBC News. Online: https://tinyurl.com/cj4nmsar
Tenbarge, K., & Chan, M. (2023. november 23.). *For teen girls victimized by 'deepfake' nudes, there is little recourse*. NBC News. Online: https://tinyurl.com/4smfxkwu
The Dalí Museum (2019. január 23.). *Dalí Lives: Museum Brings Artist Back to Life with AI*. Online: https://tinyurl.com/55fs7dv8
Trish, B. A. (2024. október 16.). *4 ways AI can be used and abused in the 2024 election, from deepfakes to foreign interference*. The Conversation. Online: https://tinyurl.com/adkry95m
Truby, J., Brown, R. D., Ibrahim, I. A., & Parellada, O. C. (2022). A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications. *European Journal of Risk Regulation*, 13(2), 270–294. https://doi.org/10.1017/err.2021.52
Tsui, M. (2022. július 25.). *Deepfake Video Examples and How to Detect One*. ExpressVPN Blog. Online: https://www.expressvpn.com/blog/how-to-spot-a-deepfake-video/
UNICEF (2024). *Cyberbullying: What is it and how to stop it*. Online: https://bit.ly/4g54qAD
US Senate (2024. június 12.). *Senate Republican Blocks Durbin's Attempt to Tackle Nonconsensual, Sexually-Explicit Deepfakes*. United States Senate Committee on the Judiciary. Online: https://tinyurl.com/2p9mz8ur
Vella, V. (2021. március 12.). *Bucks County woman created 'deepfake' videos to harass rivals on her daughter's cheerleading squad, DA says*. The Philadelphia Inquirer. Online: https://tinyurl.com/5enzzku2
Ventura, J. (2024. november 1.). *Raffensperger asks X to take down „false" video purporting to show voter fraud*. The Hill. Online: https://tinyurl.com/bdzm2eju
Wakefield, J. (2022. március 18.). *Deepfake presidents used in Russia-Ukraine war*. BBC. Online: https://www.bbc.com/news/technology-60780142
Weimann, G., & Masri, N. (2023). Research Note: Spreading Hate on TikTok. *Studies in Conflict & Terrorism*, 46(5), 752–765. https://doi.org/10.1080/1057610X.2020.1780027
Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52. https://doi.org/10.22215/timreview/1282
Winick, E. (2018. október 16.). *How acting as Carrie Fisher's puppet made a career for Rogue One's Princess Leia*. MIT Technology Review. Online: https://tinyurl.com/bdz4p7ky
Zinski, D. (2020. augusztus 18.). *Harrison Ford Is Young Han In Solo Deepfake Video*. ScreenRant. Online: https://tinyurl.com/38mrp3hk