infojegyzet

ORSZÁGGYŰLÉS HIVATALA
KÖZGYŰJTEMÉNYI ÉS KÖZMŰVELŐDÉSI IGAZGATÓSÁG
KÉPVISELŐI INFORMÁCIÓS SZOLGÁLAT

2024/20.
2024. június 13.

DEEPFAKE, AZAZ MÉLYHAMISÍTÁS: TECHNOLÓGIA ÉS JOG

A Képviselői Információs Szolgálat Infojegyzete célul tűzi a deepfake azaz mélyhamisítás technológiájának rövid ismertetését, vázlatosan áttekinti annak hatásait, valamint számbaveszi az ellene való fellépés jogi kereteit. A deepfake a dezinformáció egyik eszköze, amelyet a megtévesztés szándékával terjesztenek.

- **Mélyhamisítvány: mesterséges intelligencia által generált vagy manipulált kép-, hang- vagy videotartalom, amely létező személyekre, tárgyakra, helyekre vagy más entitásokra vagy eseményekre hasonlít, és egy személy számára hamisan hitelesnek vagy igaznak tűnhet (MI-rendelet).**
- **A deepfake egy olyan technológia, amely a mélytanulás erejét használja a hang- illetve audiovizuális tartalmak megtévesztő előállítására.**
- **A gépi tanulás a mesterséges intelligencia egy olyan alkalmazása, ahol a számítógépek az adatok felhasználásával automatikusan maguk is fejlődnek.**
- **A "mesterséges intelligencia által generált szintetikus média" kifejezést azért kezdik el használni a deepfake helyett, hogy elfedje a deepfake-hez kapcsolódó negatív konnotációt.**
- **A technológia alkalmazásának pozitív hatásai láthatóak az egészségügyben, a tanításban, a szórakoztató- és filmiparban (pl. kaszkadőr jelenetekben élő személyek helyettesítésére is), a múzeumokban és a tárlatokkal való közönség interakciók elősegítésében és általában a tartalom előállítás más területein, ide sorolva akár a marketinget is.**

Bevezetés

A deepfake, vagy mélyhamisítás, eredetét tekintve 2017 decemberére tehető; amikor egy Reditt felhasználó azt megalkotta a pornográf tevékenységet folytató híresség ek ábrázoló videók kapcsán (Thombre 2021). **A deepfake olyan digitális manipuláció,** amellyel kép- illetve hangfelvétel tartalmakat állítanak elő a mesterséges intelligenciát alkalmazó technológia segítségével a valóságtól eltérő tartalommal és valamely szándékolt hatás elérése érdekében (Gosztonyi 2024). A deepfake a személy plágiumaként is értelmezhető (Aczél 2023).

A technológia alkalmazása – bár számos pozitív lehetőséget rejthet (Gaur–Ratta 2022) –, gyakrabban jelenik meg a veszélynarratíva felől közelítve a médiában és ezzel hatással lehet a technológia felhasználási szokásainak alakítására, hívja fel e tényre a figyelmet Mezriczky Marcell (Mezriczky 2023).

Egy eseményről készült audio- és vizuális felvételt az eseményről szóló hiteles beszámolóként kezelünk. A fényképek és videók fontos információforrásként szolgálnak a mindennapi életünkben, de azon túl is kiemelkedő a jelentőségük a rendőri munkában vagy a bírósági bizonyításban. 2019-ben vált ismertté az első, deepfake hanggal szimulált csalás (Index 2019). A szakértők azóta aggódnak, hogy a rosszindulatú mélyhamisításokat politikai célokra használhatják fel, amelyek a közösségi média révén jóval gyorsabban terjedhetnek (lásd pl. a Zelenszkij elnök megadásra felhívó deepfake beszédét). Többek között ez, valamint az európai parlamenti választások is motiválták az európai pártok önkéntes magatartási kódexének elfogadását (IDEA–European Commission 2024), amelyben arra tesznek ígéretet, hogy nem készítenek, nem használnak és nem terjesztenek "*semmilyen formában megtévesztő tartalmakat*".

A deepfake technológia mögött két alapvető technológiai fejlesztés áll, az ún. mélytanulás és a generatív adverzális hálózatok. A mélytanulás a gépi tanulás egy fajtája. Lényege, hogy a számítógép ún. neurális hálózatok segítségével elemzi az adathalmazokat, abból a célból, hogy mintákat találjon. A neurális hálózatok az agyunk működését utánozzák. A hatékonyság záloga a megadott adatok milyensége. Ezért a jó mélytanulási rendszerhez elengedhetetlen a nagy betanító adathalmaz. A rendelkezésre álló adatok minőségétől és az algoritmus által használt tényezőktől függően az eredmény többé-kevésbé reális lehet.

Generatív adverzális hálózatok (GAN)

A deepfake technológia minőségében és hozzáférhetőségében nagy ugrást jelentett a generatív adverzális hálózatok (GAN) adaptálása, amelyet Ian J. Goodfellow et al. javasolt 2014-ben. A GAN két egymással versengő modellel dolgozik: egy ún. generatív és egy másik, ún. diszkriminatív modellel. A generatív modell a rendelkezésre álló képzési adatok alapján tartalmat hoz létre. Az adatokat elemezve olyan tartalmat állít elő, amely a képzési adatokban szereplő példákat a lehető legjobban utánozza. Ezt követően, a tesztelést a diszkriminatív modell végzi. A generatív modell által létrehozott eredményt aszerint méri fel, hogy vajon mekkora annak a valószínűsége, hogy a tesztelt minta nem a generatív modellből, hanem az adathalmazból származik.

A mélyhamisítások létrehozására szolgáló **képzési adatok** különböző módon alkalmazhatók a videó- és képmélyhamisítások esetében (Masood 2022):

- Arccsere: célja, hogy egy személy arcát cserélje a videón szereplő személy arcának helyére;
- Attribútum-szerkesztés: a videóban szereplő személy jellemzőit kívánja megváltoztatni, pl. stílus vagy hajszín változtatással;
- Arcok mimikai újrajátszása: az arckifejezések átvitele egy személy arcáról a célvideóban szereplő személyre;
- Teljesen szintetikus anyag: valódi anyagot használnak például az emberek kinézetének betanításához, de a kapott kép teljesen kitalált, nem létező személyt jelenít meg.

A technológia egyes kockázatai

A történelem során az új kommunikációs technológiák megjelenését, elterjedését és használatát kísérő jelenség az ún. technopánik (Thierer 2013). A pánik narratíva olyan fogalmakkal is társul, mint az "információs apokalipszis" (Paris–Donovan 2019) vagy "valóságapátia" (Vincent–Gismondi 2021), amely azt a társadalmi jelenséget írja le, hogy a deepfake hamisítások növekvő mennyisége aláássa a demokrácia működésébe, a hatóságokba és a hivatalos tényekbe vetett bizalmat is (Thombre 2021), végső soron olyan helyzethez vezethet, amelyben a polgároknak már nincs közös valóságuk, vagy okozhat társadalmi zavart azzal kapcsolatban, hogy mely információforrásokat tekinthetnek megbízhatónak.

Az új kommunikációs technológiáknak számos előnye is ismert. Azonban ami a jogszerű felhasználás oldalán előnyt jelent, a másik oldalon is előny lehet. Az 5G technológia által kínált többlet sávszélesség például lehetővé teheti a felhasználók számára, hogy a felhőalapú számítástechnika erejét használva a videófolyamokat akár valós időben manipulálják. A deepfake-technológiák ezért videokonferencia-környezetben, élő közvetítésű videoszolgáltatásokban és a televíziózásban is alkalmazhatók lehetnek (Europol 2020).

A bűnözés, mint szolgáltatás (’crime as a service’, vagy angol rövidítéssel ’CaaS’) a technológiákkal párhuzamban várhatóan fejlődni fog, és megjelenhet a technológiák és eszközök, valamint a kiberalapú bűnözés elősegítéséhez szükséges tudás áruba bocsátása révén (Europol 2020).

A deepfake technológia segítségével elkövethető **bűncselekmények** közé tartoznak például a(z) (Európai Parlament 2021):

- zaklatás, egyének online megalázása;
- zsarolás, csalás;
- okirathamisítás, annak elősegítése;
- online személyazonosság meghamisítása;
- nem konszenzuális pornográfia;
- gyermekekkel szembeni online szexuális zaklatás vagy kizsákmányolás;
- büntető nyomozás elektronikus bizonyítékainak meghamisítása vagy manipulálása;
- pénzügyi piacok megzavarása;
- dezinformáció terjesztése, közvélemény befolyásolása, nyilvánosság manipulálása;
- szélsőséges vagy terrorista csoportok narratíváinak támogatása;
- társadalmi nyugtalanság vagy politikai polarizáció szítása.

Válaszok a deepfake jelenségre

A mélyhamisítás olyan technológiai kihívás, amelyre részben a technológia adhat választ is. Rana et al. (2022) például összegzi a mélyhamisítás felismerésével foglalkozó kutatásokat a 2018–2020 közötti időszakból és 4 csoportba sorolja az alkalmazott technológiai módszereket: mélytanulás-alapú technikák, klasszikus gépi tanulás-alapú módszerek, statisztikai technikák és blokklánc-alapú technikák.

A bűnüldöző szervek munkájában is várhatóak változások, ez kihathat nemcsak az audiovizuális anyagok létrehozására, de tárolására, védelmére és elemzésére is. Érdemes már kipróbált és máshol bevált módszereket alkalmazni az audiovizuális felvételek készítésekor, pl. bizonyos beállítások tanúsítása a bírósági felhasználáshoz, valamint szükség lehet hamisítás elleni technikai és szervezési biztosítékok alkalmazására, hogy bizonyítani lehessen a felvételek hitelességét.

Az igazságszolgáltatás számára is fontossá válhat a felvételek keresztellenőrzése. A bírósági eljárásban az audiovizuális bizonyítékok kiemelkedő jelentőségűek, azok hitelességében általában bíznak. Akár a gyanúsított telefonjáról vették ki a fájlt, akár a közösségi médiából töltötték le, akár a bűncselekmény helyszínéhez közeli CCTV-rendszeréről kapták, az ábrázolt jelenet hitelességét általában nem kérdőjelezik meg. A mélyhamisítások terjedésével azonban egyre fontosabbá válik a tartalmak keresztellenőrzése, amely más típusú szaktudást, szakértői munkát, illetve megfelelő szakértők bevonását feltételezi (Europol 2022).

A deepfake terjedésének leggyorsabb eszköze a **közösségi platform**. Egyes platformok pl. a Meta 2020 elején új politikát jelentett be, amely megtiltja a mélyhamisítványok használatát a platformjain, azaz eltávolítja onnan az AI által szerkesztett olyan tartalmakat, amelyek nagy valószínűséggel félrevezetik az embereket (Europol 2022). A Meta 2024 áprilisában újabb változásokat tett közzé; a videó-, hang- és képtartalmak szélesebb körét fogja a "*Made with AI*", azaz MI-vel készült, jelzővel ellátni akkor, ha az iparágban elterjedt, mesterséges intelligenciával előállított képekre utaló szabványos képjelzőket észleli, vagy ha az emberek közzéteszik, hogy mesterséges intelligencia által generált tartalmat töltenek fel (Techcrunch 2024. ápr. 5.). A Meta már most is hozzáadja a "*Képzelt MI*" feliratot az MI funkcióval létrehozott fotórealisztikus képekhez.

A technológiai nagyvállalatok, mint a Twitter és a Meta szabályozása a deepfake technológia vonatkozásában nagymértékben befolyásolhatja az emberek reakcióját. Mindezek mellett azonban számos kutatás utal arra, hogy **a médiaműveltségre való nevelés** csökkentheti a dezinformációs üzenetek hatását (Hwang et al. 2021). Éppen ezért az **Európai Bizottság** több módon is hozzájárul a médiaműveltség és dezinformáció elleni küzdelemhez, pl. egy zárójelentés készült a dezinformáció kezelése és a digitális írástudás előmozdítása oktatással és képzéssel címmel (2022), illetve elkészült az Iránymutatások tanárok és oktatók számára a dezinformáció elleni küzdelemről és a digitális írástudás előmozdításáról című kiadvány (2022) is.

Deepfake a jog tükrében

A deepfake maga egy eszköz és nem a tartalom, azonban a technológia segítségével előállított tartalom a szólásszabadságon túlmutatva eshet jogi korlátozás alá; felvethet pl. adatvédelmi (Eszteri 2023), személyiségi jogi vagy büntetőjogi aggályokat (Ambrus 2021, Sorbán 2020).

Amerikában a deepfake elleni egységes fellépés gátja az alkotmány első kiegészítésével védett szólásszabadság (O'Donnell 2021). Egyes tagállami gyakorlatok pl. Kalifornia és Texas államban a választások tisztaságának érdekében fellépést engednek a politikai jelölteknek kárt okozó deepfake videókkal szemben. New York és Nebraska államok kifejezetten tiltják a rossz szándékkal létrehozott és terjesztett deepfake tartalmakat (Lendvai 2023).

Kína 2022-ben jogszabályi tiltást fogadott el, amelynek alapján "*Semmilyen szervezet vagy magánszemély nem használhatja a mélyszintézist nemzetbiztonságot sértő, a társadalmi stabilitást romboló, a társadalmi rendet felborító,*

mások törvényes jogait és érdekeit sértő vagy más, törvények és rendeletek által tiltott tevékenységekre".

Az **Európai Unió** szabályozási környezete igen sokrétű, magában foglal uniós és tagállami szintű kemény és puha szabályokat. A két legutóbbi a témát is érintő jogszabály: a Mesterséges Intelligenciáról (MI) szóló rendelet (Regulation (EU) 2024/…), amelyet mind az Európai Parlament és az Európai Unió Tanácsa jóváhagyott, hivatalos közzététele júniusban várható; valamint a Digitális szolgáltatásokról szóló rendelet (EU 2022/2065). Az MI rendelet meghatározza a deepfake fogalmát és az átláthatóság érdekében előírja e rendszerek szolgáltatói és felhasználói számára az ilyen rendszerek, tartalmak megjelölését. Az utóbbi, DSA rendelet, a platformszabályozás eszköze, célja olyan intézkedések kikényszerítése, amelyek biztosítják a jogellenes tartalmak, így pl. a deepfake szolgáltatói szűrését (Miklós 2023).

A magyar szabályozásban a deepfake kifejezetten nem szerepel, azonban a Büntető Törvénykönyv (Btk.) releváns rendelkezései alkalmazhatók a mélyhamisítással elkövetett bűncselekményekre is (pl. zaklatás, becsületsértés, szexuális zsarolás, rémhírterjesztés). Egyes kutatók szerint azonban (Mickolczi–Szatmáry 2018), a Btk. tényállásai a deepfake vonatkozásában nem adnak arányos válaszokat.

Az egyéni és társadalmi károk negatív hatásainak kezelésére nem elegendő a jog eszköze, amely inkább reaktív, hiszen a kár sokszor már a közzététellel bekövetkezik. Ezért fontos cél e károk megelőzése, azaz a deepfake technológia nemkívánatos felhasználásainak elkerülése és a megakadályozásra irányuló intézkedések ösztönzése.

Források:

- Aczél Petra–Veszelszki Ágnes (szerk.) (2023): Deepfake: A valótlan valóság. Gondolat Kiadó
- Aczél Petra (2023): A deepfake mint hazugság: együttműködés a megtévesztésben. In: Aczél–Veszelszki
- Ambrus István (2021): Digitalizáció és büntetőjog. Wolters Kluwer. Új jogtár verzió
- Delfino, R (2024): Pay-to-play: Access to Justice in the Era of AI and Deepfakes. Loyola Law School, Los Angeles Legal Studies Research Paper No. 2024-08.
- Eszteri Dániel (2023): A deepfake-technológia adatvédelmi értékelése a GDPR tükrében. In: Aczél–Veszelszki
- Gosztonyi Gergely–Lendvai Gergely (2024): Deepfake és dezinformáció Mit tehet a jog a mélyhamisítással készített álhírek ellen? Médiakutató, 2024. tavasz XXV. évf. 1. szám, 41–49. o.
- Köbis et al. (2021): Fooled twice: People cannot detect deepfakes but think they can. iScience Volume 24, Issue 11.
- Lendvai Gergely Ferenc (2023): Deepfake a szólásszabadság tükrében – reflexiók a jog perspektívájából. In: Aczél–Veszelszki
- Masood et al. (2022): Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward. Applied Intelligence Volume 53, pages 3974–4026
- Mezriczky Marcell (2023): Ne higgy a szemének! A deepfake online sajtóreprezentációja 2018 és 2022 között. In: Aczél–Veszelszki
- Miklós Gellért (2023): A deepfake-tartalmak szabályozása az Európai Unió jogában. In: Aczél–Veszelszki
- Miskolczi B.–Szathmáry Z. (2018): Büntetőjogi kérdések az információk korában. HvgOrac kft.
- Thombre, M. (2021): Deconstructing Deepfake: Tracking Legal Implications and Challenges. International Journal of Law Management & Humanities, 4, 2267-2274.

Készítette: Dr. Szalay Klára
Képviselői Információs Szolgálat
E-mail: infoszolg@parlament.hu

infoszolg

Internet: www.parlament.hu/infoszolg
Intranet: intra.parlament.hu/infoszolg/
Telefon: (1) 441-6486