CTI elemzés

Szórakozástól a jogsértésig

– hogyan használható a DeepFake

1996

NEMZETBIZTONSÁGI SZAKSZOLGÁLAT

NEMZETI KIBERBIZTONSÁGI INTÉZET

Tartalomjegyzék

Bevezetés

A mesterséges intelligencia (MI) rohamszerű fejlődésének köszönhetően 2026-ra jelentős mértékben átalakultak a digitális tartalmak előállítására és fogyasztására vonatkozó szokások.

A DeepFake technológia folyamatos térnyerése lehetővé teszi, hogy felhasználók széles köre a szándékuktól függetlenül olyan képi-, videó- és/vagy hangfelvételeket állítsanak elő, amelyek a valóságban sosem történtek meg vagy hangoztak el. Ráadásul az idő előrehaladtával nemcsak egyre több platform áll rendelkezésre az MI segítségével létrehozható tartalmak előállításához, hanem egyre alaposabb, valósághűbb tartalmak is készülnek. Emiatt jelentősen megnehezedik annak megállapítása, hogy a felhasználó MI által létrehozott tartalommal találkozik-e vagy sem. Éppen ezért az egyik legfontosabb kérdés, amit 2026-ban érdemes feltenni a digitális tartalomfogyasztás kapcsán az, hogy:

„Elhihetjük, amit a neten látunk és hallunk?”

Persze a kérdést nem érdemes kizárólag az internetre korlátozni, ugyanis számos terület alkalmaz manapság DeepFake technológiát. Legális oldalon például ott vannak az MI által generált óriásplakátok, televíziós és nyomtatott reklámanyagok, illegális oldalon pedig a különböző kibercsalási módszerek, mint például a DeepFake telefonhívások. Ezekről és más felhasználási területekről bővebben lesz még szó az elemzés „Felhasználási területek” című fejezetében.

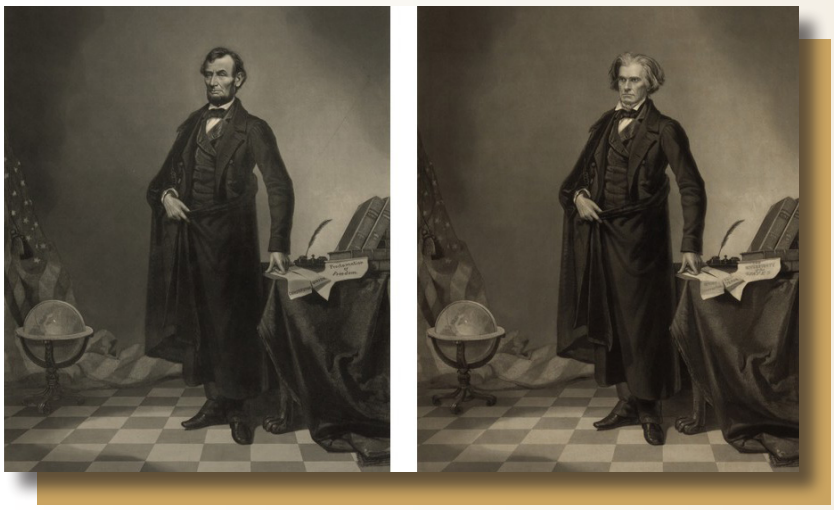
A fenti kérdés, illetve a technológia felhasználási céljának bizonytalanságából adódóan, vagyis, hogy jó vagy rossz célra használják a DeepFake tartalmakat, komoly bizalmi problémákat vet fel a digitális tartalomfogyasztás terén, ami hosszú távon a felhasználók teljes bizalomvesztéséhez vezethet.

Jelen dokumentum célja – amellett, hogy választ találjon a fenti kérdésre –, hogy új szemszögből mutassa be a DeepFake technológiát, amely során elemzésre kerül, hogy mely területekre férkőzött be az MI alapú tartalomkészítés, milyen szabályozási módszerekkel igyekeznek kordában tartani azt, illetve tanulmányozásra kerülnek az elmúlt évek tapasztalatai is.

Történelem és technológia

Ahhoz, hogy megértsük a DeepFake tartalmakat és az ezekkel összefüggésben felmerült kérdéseket, érdemes áttekinteni a technológia működését és a fejlődése mögötti motivációt.

Mivel a DeepFake egy viszonylag új technológia, a szó mai értelmében még nem létezett a 19-20. században, viszont a valóság képi manipulációjának igénye már ebben a korban is megjelent. A 19. században, a fotográfia korai időszakában már előfordult, hogy különböző módszerekkel, például több negatív összeillesztésével vagy kézi retusálással manipulálták a képeket. Egyik ilyen ismert eset Abraham Lincoln portréja, ahol valaki más testére illesztették az elnök fejét.



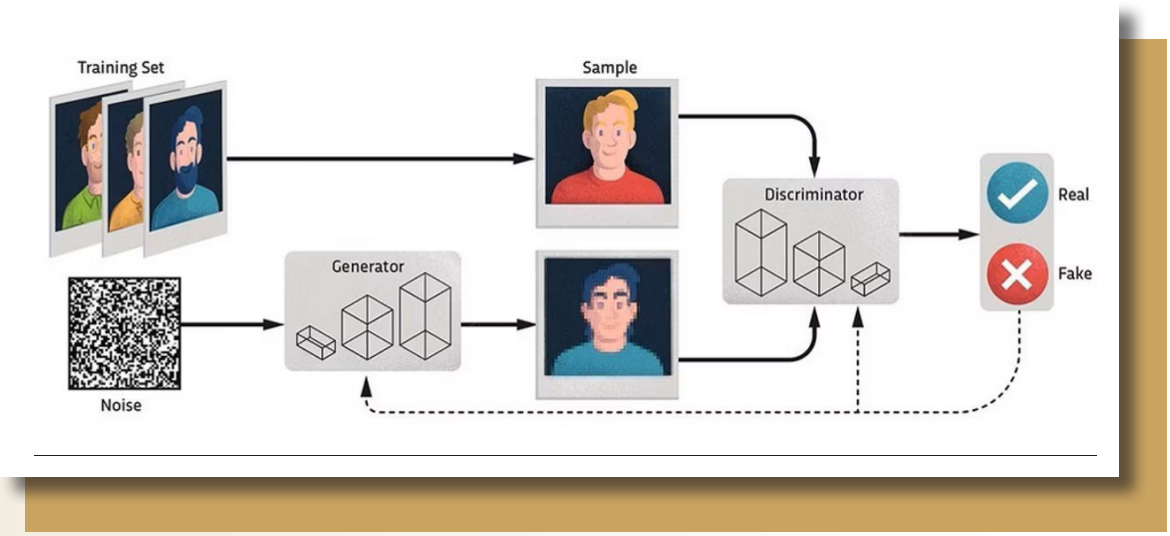
1. kép Abraham Lincoln portréja

(kép forrása: nationalgeographic.com)

A 20. században már politikai propagandához is felhasználták a képek manipulálását, például Joseph Stalin rendszere alatt embereket retusáltak ki képekről, ami gyakorlatilag a történelem átírását jelentette.

A neurális hálózatok és a gépi tanulás fejlődése jelentik a DeepFake alapjait, különösen a képfelismerési és generálási területeken. A 2010-es évek elején jelentek meg azok a módszerek, amelyek képesek voltak már arcok felismerésére és manipulálására.

2014-ben Ian Goodfellow és kutatótársai fejlesztették ki a mai DeepFake mögött álló Generative Adversarial Network (GAN) technológiát, egy olyan gépi tanulási modellt, ami képes új adatokat létrehozni a már betanított adathalmazokból. A GAN két neurális hálózatból (Generátor és Diszkriminátor) áll, amelyek folyamatosan „versenyeznek” egymással, hogy minél élethűbb tartalmat tudjanak előállítani. Míg a Generátor mesterséges tartalmat hoz létre, addig a Diszkriminátor megpróbálja eldönteni, hogy az adott tartalom valódi-e vagy hamis. Ez a folyamat végül egyre valósághűbb tartalom létrehozásához vezet.



2. kép GAN technológia
(forrás: ibm.com)

Néhány évvel később, 2017-ben jelent meg a DeepFake kifejezés egy Reddit felhasználó által, aki a GAN technológia segítségével más emberek képeinek felhasználásával, felnőtt videótartalmakat állított elő és tett közzé az interneten. Bár akkoriban az ilyen és ehhez hasonló videók alacsony minőségűek voltak, a DeepFake elterjedésével és a technológia finomhangolásával egyre jobb minőségben készülnek képi- és hangalapú DeepFake tartalmak.

2018-ban a DeepFake bekerült a köztudatba, hála a könnyen hozzáférhető, nyílt forráskódú eszközöknek. Azóta folyamatosan egyre könnyebbé válik a valósághű tartalmak előállítása.

2023-ban a DeepFake eszközök piaca robbanásszerű növekedést mutatott, közel 44%-kal nőtt az ilyen jellegű fejlesztések száma. Sajnos azonban a technológia elterjedését nagy mértékben ösztönözték az általában nőkről készült, beleegyezés nélküli, gyakran explicit tartalmak készítése. A Security Hero jelentése szerint 2023-ban az online DeepFake videók körülbelül 98%-a ilyen jellegű volt, és ezeknek csupán 1%-ában voltak férfiak a célpontok.

A technológia fejlődésének ugrásszerű növekedését tehát nem kimondottan a „jó szándék" okozta, viszont időközben számos alkalmazási területre beférkőzött, ahol hasznosnak bizonyult, ezek az elemzés „Felhasználási területek" című fejezetében részletesen ismertetésre kerülnek.

Múltbéli becslések

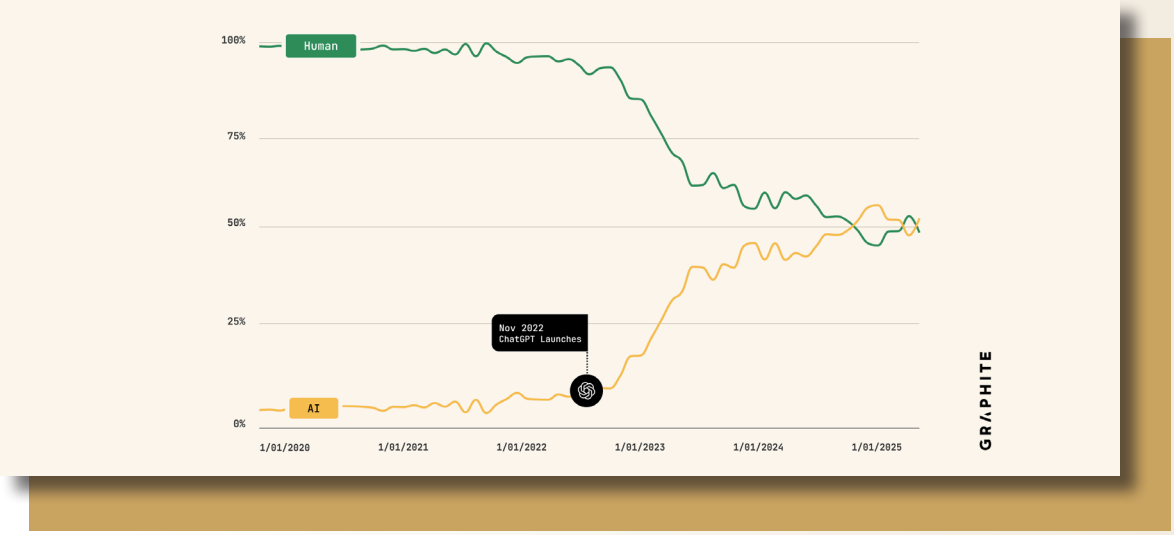
Az NBSZ NKI által egy, még 2022-ben kiadott „Mekkora fenyegetést jelentenek valójában a deepfake-ek?" című kiberbiztonsági elemzése szerint a szakértők akkoriban úgy vélték, hogy 2026-ra az online tartalmak 90%-át már mesterséges intelligencia fogja előállítani.

> *Megjegyzés: Ezen a ponton érdemes tisztázni, hogy milyen internetes tartalmakról lehet itt szó, illetve mit jelent a szintetikus média kifejezés. Szintetikus médiának tekintünk minden olyan interneten elérhető médiatartalmat, például közösségi média posztot, hirdetést, blogbejegyzést stb., aminek az előállításához mesterséges intelligenciát használtak. Ez lehet szöveg, kép, videó hang és ezek tetszőleges variációi.*

Sajnos nem áll rendelkezésre egy egységes, mindenki által elfogadott szám arra vonatkozóan, hogy az interneten található tartalmak hány százalékát állították elő mesterséges intelligenciával, viszont több becslés és kutatás is igyekezett már meghatározni ennek mértékét.

Egyik ilyen a Graphite nevű cég egy viszonylag friss felmérése, amelynek eredményei azt mutatják, hogy a vizsgálatban résztvevő internetes tartalmak több mint fele került előállításra mesterséges intelligenciával. A felmérést közel 65 ezer 2020 és 2025 között közzétett URL-en végezték. Az eredmények jól mutatják, hogy a ChatGPT 2022-es bevezetése ugrásszerűen megemelte az MI generált cikkek számát, viszont az elmúlt 12 hónapban jól láthatóan stabilizálódott ennek mértéke. Ennek egyik lehetséges magyarázata, hogy a szakemberek tapasztalatai szerint például az MI generált cikkek nem teljesítenek jól az internetes keresések során.

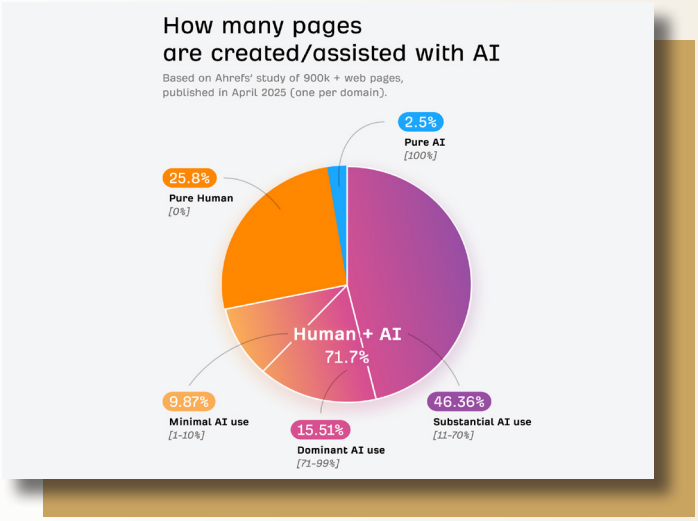
Megjegyzés: A Graphite kimondottan a megjelent szöveges publikációkat vizsgálta felmérésében, így más szintetikus médiatartalomra vonatkozó adat, például az MI generált képek és videók tekintetében nem vontható le következtetés a megoszlásra vonatkozóan.



3. kép Emberek és MI által létrehozott internetes tartalmak megoszlása

(Forrás: graphite.io)

Az Ahrefs blog felmérése ennél magasabb arányt mutatott, az általuk elemzett 900.000 weboldal vonatkozásában, ugyanis az új weboldalak közel 74%-a tartalmazott MI által létrehozott részeket.



4. kép Emberi és MI által létrehozott weboldalak megoszlása

(Forrás: ahrefs.com)

A DeepFake szabályozása

Nincs kimondottan, kizárólag csak a DeepFake szabályozására vonatkozó törvény, de több olyan jogszabály is rendelkezésre áll, amelyek foglalkoznak a technológiával összefüggő jogi kérdésekkel.

Egyik ilyen az Európai Unió mesterséges intelligenciáról szóló rendelete (AI Act vagy MI rendelet), amelyben hivatalosan definiálásra kerül a DeepFake, mint MI által generált vagy manipulált valósnak tűnő tartalom. A rendelet átláthatósági kötelezettséget ír elő a DeepFake tartalmakra, tehát egyértelműen és megkülönböztethetően fel kell tüntetni, hogy a tartalom mesterségesen került létrehozásra vagy manipulálásra. A rendelet arra is kitér, hogy mindez nem akadályozhatja a szabad véleménynyilvánításhoz, valamint a művészet és a tudomány szabadságához való, valamint a Chartában garantált jogokat.

Magyarországon külön törvény (2025. évi LXXV. törvény) jelent meg az Európai Unió mesterséges intelligenciáról szóló rendeletének magyarországi végrehajtásáról, amelyben külön intézményrendszer került létrehozásra az MI rendszerek jogszerű alkalmazásának felülvizsgálatára.

Ezen kívül bár Magyarországon nincs külön DeepFake törvény, a Büntető Törvénykönyv (Btk.) szerint a DeepFake tartalmak illegálisnak tekinthetők, ha azokat rágalmazáshoz, becsületsértéshez, zaklatáshoz, szexuális visszaélésekhez vagy álhírek terjesztésére használják fel.

Ezek figyelembevételével megállapítható, hogy önmagukban nem minősülnek illegálisnak a DeepFake tartalmak, azok felhasználási módja határozza meg, hogy jogsértők-e vagy sem.

Nemzetközi kitekintés

Az Amerikai Egyesült Államok 2025 májusában fogadta el a TAKE IT DOWN törvényt (Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act), amely bűncselekménnyé minősíti a nem beleegyezésen alapuló – valódi vagy mesterséges intelligenciával létrehozott vagy manipulált – intim képek és videók közzétételét és terjesztését, továbbá előírja a platform üzemeltetői számára, hogy bejelentéstől számítva 48 órán belül el kell távolítaniuk az ilyen jellegű tartalmakat.

Az Egyesült Királyságban 2023 októberében lépett hatályba az Online Safety Act, amelynek célja az online tér, különösen a közösségi médiaoldalak biztonságosabbá tétele és szabályozása. A törvény lényege, hogy minél hatékonyabban csökkentsék a káros tartalmakat azok visszaszorításával, illetve arra ösztönözni a platformok üzemeltetőit, hogy minél gyorsabban eltávolítsák azokat.

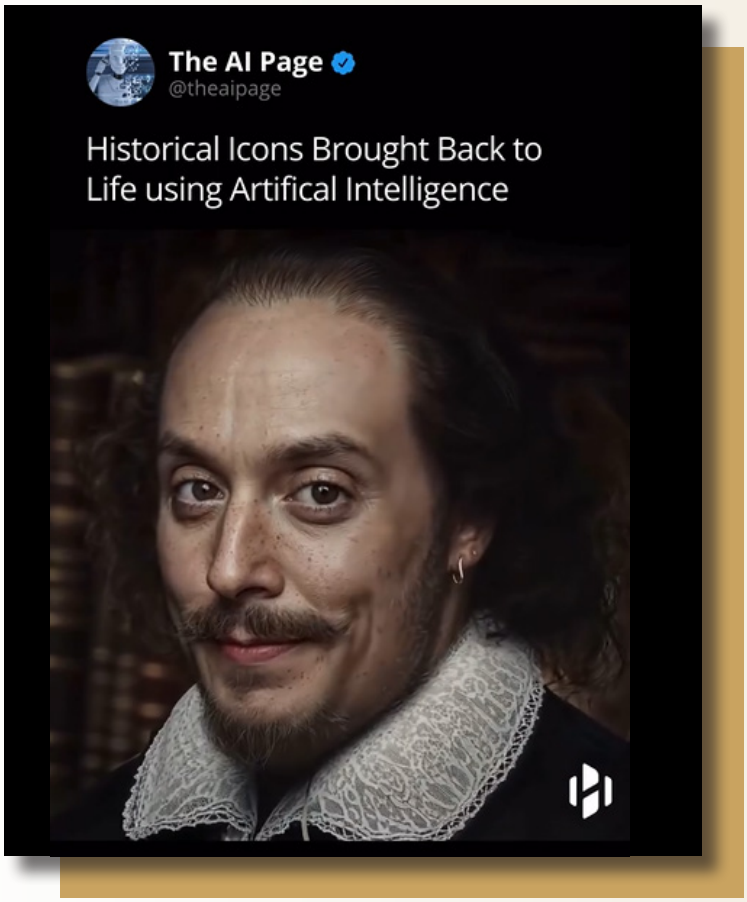
A kínai Provisions on the Administration of Deep Synthesis of Internet Information Services (2024) című rendelet 2023 januárjában lépett hatályba, és meglehetősen hasonlít az uniós MI rendelethez. A kínai rendelet célja a DeepFake technológia szabályozása és kontrollálása, valamint a felhasználói jogok védelme. Emellett jelentős hangsúlyt fektet az MI etikus és társadalmilag megfelelő használatára.

Felhasználási területek

Az előző fejezetben tett megállapítás tükrében, érdemes külön választani a legális és illegális felhasználási területeket. A szórakoztató- és fogyasztóiparban, az oktatásban, illetve még terápiás célból is jogszerűen felhasználható a DeepFake technológia.

- Filmek és különböző médiaanyagok (pl.: szinkron) készítésekor a DeepFake gyorsabbá, egyszerűbbé és költséghatékonyabbá teheti a speciális effektek használatát, nem mellesleg a kaszkadőrjelenetekhez is biztonságosabb megoldást jelenthet. Fontos megjegyezni, hogy a szórakoztatóiparban használt DeepFake technológia használata kizárólag a színészek beleegyezésével tekinthető legitimnek.
- Marketing és reklámtevékenységekhez használt generatív MI használata jelentősen egyszerűsíti a hirdetéskészítési folyamatokat, ráadásul így ugyanannyi vagy még több reklámanyag elkészítése kevesebb időt és energiát vesz igénybe. Emellett a demográfiai, illetve a felhasználók online viselkedésére és böngészésére vonatkozó adatok ilyen jellegű elemzése jelentős hatást gyakorolt a személyre szabott hirdetési kampányokra is. A hirdetőknek ilyen esetekben az átláthatósági követelményeknek való megfelelés mellett, a felhasználók egyértelmű beleegyezésére is szükségük van a jogszerű eljáráshoz.
- Oktatás és ismeretterjesztés céljából is lehet DeepFake technológiát alkalmazni, például interaktív tananyagok létrehozásához (pl.: a történelmi személyek „maguk” is elmesélhetik a tananyagot). Ilyen esetekben is fel kell tüntetni, hogy az adott tananyag MI által lett létrehozva.

A TheAIPage Instagram oldalán közzétett videóban például történelmi alakok fényképei kerültek „retusálásra” a mesterséges intelligencia segítségével, aminek köszönhetően nemcsak azt láthatjuk, hogyan nézhettek ki a fényképük alapján az életben, hanem a legtöbben még mozogni és mosolyogni is képesek a videóban.

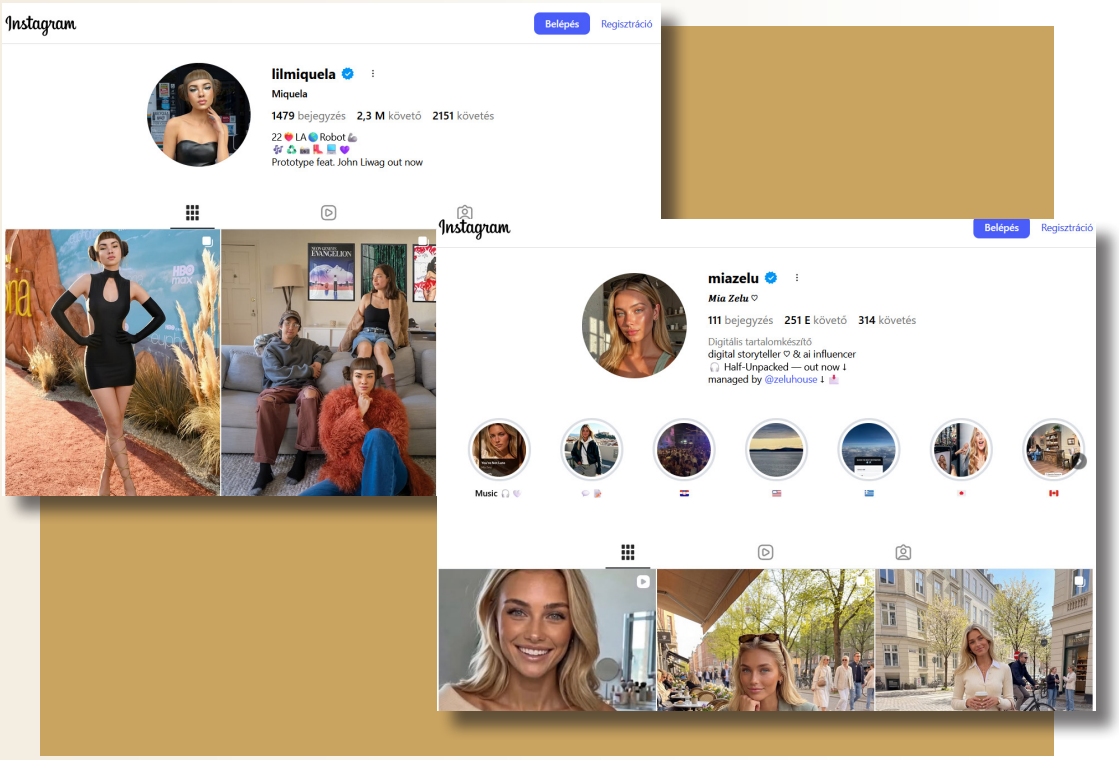


5. kép Részlet a TheAIPage videójából

(Forrás: @theaipage)

- Játékok és digitális tartalmak készítésében is egyre népszerűbb technológiának bizonyult a DeepFake, ugyanis amellett, hogy realisztikusabb karakterek hozhatók létre általa, új iparág is megjelent a digitális tartalomgyártásban, méghozzá az MI-influenszerek, vagyis a mesterséges intelligencia által létrehozott digitális személyiségek formájában.

Ilyen például a 19 éves Lil Miquela néven ismert, több mint 2 millió követőbázissal rendelkező virtuális közösségi média személyiség és énekes, valamint Mia Zelu digitális tartalomkészítő.



6. kép Lil Miquela és Mia Zulu Instagram fiókjai

(Forrás: Instagram.com [1], [2])

- Jelenleg ugyan még jogi és etikai szempontból vizsgálják, de kísérleti jelleggel már felmerült a DeepFake technológia traumaterápiás használata. A lehetséges alkalmazási területek közé egyelőre a gyászfeldolgozás és a szexuális erőszakkal kapcsolatos traumák kezelése tartozik. A National Library of Medicine kutatásában például a szexuális erőszakkal kapcsolatos PTSD (poszttraumás stressz szindróma) kezelése során az áldozat Zoom-on keresztül beszélgethetett az elkövetőjét megszemélyesítő terapeutával. A kísérlet lényege az volt, hogy az „elkövető” empatikus válaszaival és viselkedésével segítse az áldozatot az önmegbocsátásban azáltal, hogy felhagy az önhibáztatással.

A technológia megtévesztő erejére természetesen a kiberbűnözők is felfigyeltek, így nem meglepő, hogy egyre több támadási kampány egészül ki vagy alapszik teljes egészében a képek és hangok hamisítását célzó módszerekre. Általánosságban elmondható, hogy minden DeepFake technológián alapuló visszaéléssel kapcsolatos tevékenység illegálisnak minősül.

- A DeepFake egyik ilyen legnépszerűbb felhasználási módja a nem beleegyezésen alapuló, általában, de nem kizárólag intim tartalom terjesztése. A deep-fake.ai szerint a DeepFake videók túlnyomó többsége – körülbelül 96%-a – nőket célzó, nem konszenzuson alapú pornográf tartalom. Jellemzően hírességekről, politikusokról és egyéb közszereplőkről készülnek ilyen jellegű videók, melynek egyik lehetséges magyarázata, hogy sok fénykép érhető el róluk a neten, viszont ettől függetlenül bárki áldozattá válhat.

  Ezen felhasználási módnak egyik talán még rosszabb alfaja, amikor kiskorúak az áldozatok.

- Egyre nagyobb teret hódítanak a DeepFake-alapú megtévesztések és pénzügyi csalások. Egyik leginkább elterjedt módszer, mikor a kiberbűnözők hitelesen leutánozzák a vállalati vezetők hangját (BEC típusú csalások) és kommunikációs stílusát, aminek reprezentálásához gyakran a nyilvánosan, például közösségi médiában elérhető vagy konferenciaszereplések során gyűjtött információkat használják fel.

  Emellett bármely más social engineering alapú támadást is meg lehet támogatni DeepFake technológiával, például az unokázós és romantikus csalásokat.

- A DeepFake jól használható dezinformációs kampányokhoz és politikai manipulációhoz. Tardi Roland PhD kutató cikkében éppen azt vizsgálja, hogyan illeszkednek a DeepFake tartalmak a post-truth politika fogalmába. A post-truth politika vagy „igazság utáni politika" egy olyan politikai közeget jelent, amelyben jelentősen háttérbe szorulnak az objektív vélemények, és inkább a szubjektív, érzelmi és személyes meggyőződéseken alapuló meglátások kerülnek előtérbe. Ebbe a megközelítésbe jól illeszkednek a digitális manipulációs eszközök, mint a DeepFake, a fake news (hamis hírek) és egyéb összeesküvés-elméleteket alátámasztó módszerek.

- A DeepFake a BEC és egyéb social engineering módszeren alapuló csalásokhoz hasonlóan, más megszemélyesítéshez és személyazonosság-lopáshoz is felhasználható, például közösségi oldalakon létrehozott profilokhoz és a biometrikus azonosítások megkerüléséhez, bár utóbbi esetén a modernebb rendszerek már képesek védekezni az ilyen jellegű megtévesztések ellen.

- A technológiát súlyos pszichés bántalmazáshoz is felhasználhatják, például rágalmazáshoz, zaklatáshoz vagy zsaroláshoz. Főleg vezető beosztásban lévő emberek vagy híresség a célpontjai a rágalmazásnak, míg a zaklatás általában az iskolás korúakat (cyberbullying) érinti, a zsarolás pedig sokszor része az előző két visszaélési formának is. Az elkövetők bárkiről készíthetnek szexuális vagy illegális tartalmat, azonban azokból a felhasználókból lesz áldozat, akik helyzetükből vagy személyiségükből adódóan félnek a tartalmak nyilvánosságra kerülésétől és hajlandók sokat tenni ennek elkerülése érdekében.

Megtörtént esetek elemzése

Híresség és politikusok

- Az egyik első, széles körben és szándékosan megtévesztő módon alkalmazott DeepFake a Volodimir Zelenszkij elnökről készült videó, amelyben arra szólítja fel az ukrán állampolgárokat, hogy adják meg magukat az orosz katonáknak. A videó nem sokkal azután került nyilvánosságra, hogy Oroszország 2022. február 24-én megtámadta Ukrajnát. Bár a videó nem volt túl jó minőségű, ráadásul Ukrajna gyorsan cáfolta a videóban elhangzottakat, az eset jól szemlélteti a technológia háborús célú dezinformációs erejét, különös tekintettel a pánikkeltésre és a kommunikáció elhiteltelenítésére egy ilyen kritikus időszakban.

- Másik régebbi eset a Taylor Swift botrány. Az énekesnőről hamisított intim képek kerültek fel az X közösségi médiaoldalra, és mivel a platform az első 17 órában nem foglalkozott az üggyel, elképesztő sebességgel terjedtek a képek. Az első ilyen posztot több mint 45 millióan tekintették meg és 42 ezren osztották meg. A rajongók közös összefogással vették fel a harcot az énekesnő becsületéért és számos „Taylor Swift AI" vagy „Taylor Swift deepfake" címszóval ellátott posztot generáltak, amelyek igyekeztek elfedni az intim képeket az azokat keresők elől, ráadásul célkeresztbe kerültek az intim képek terjesztői is. Ezt követően az X kénytelen volt rövid időre letiltani a „Taylor Swift" keresőkulcsszót. Az eset kapcsán több amerikai politikus is felszólalt, ráadásul az ügy hatására nemcsak az Egyesült Államokban, hanem Európában is megkezdődtek a DeepFake szabályozására tett jogi erőfeszítések.

- Collien Fernandesről, a német tv-s műsorvezetőről a saját exférje készített hosszú éveken át DeepFake intim videókat és terjesztette azokat a saját maga által létrehozott kamuprofilokról. Mivel a német jogrendszer nem rendelkezett a megfelelő eszközökkel, Collien a spanyol bírósághoz fordult, hogy eljárás induljon a volt férje, Christian Ulmen ellen, aki nem mellesleg szintén ismert a német médiában. Spanyolországban nemcsak szigorúbbak a törvények, de működnek külön olyan ügyészségek és bíróságok, amelyek kifejezetten a nők elleni erőszakra specializálódtak, így az ügyet nemcsak polgári perként, hanem büntetőeljárásként is kezelik. Az ügy hatására a német igazságügyi miniszter, Stefanie Hubig bejelentette, hogy a tervezettnél korábban kerül benyújtásra új törvényjavaslata, amelynek célja a nők védelmének erősítése, így kimondja, hogy akár két év börtönbüntetést is vonthat maga után a DeepFake tartalmak készítése és terjesztése.

Iskolai zaklatás DeepFake-kel

- Egy pennsylvaniai iskolában egy diák több tucat lány iskolatársáról készített és terjesztett mesterséges intelligenciával létrehozott intim fényképeket. Annak ellenére, hogy az iskola tudott az ügyről, nem tett sem bejelentést, sem a szükséges intézkedéseket a képek felszámolását illetően. Ennek köszönhetően a képek nemcsak, hogy tovább terjedtek, de még több áldozatról készültek hamis képek. Az eset kapcsán hatalmas közösségi botrány alakult ki, a szülők és a diákok együtt tüntettek az iskola ellen, ami a lemondások mellett azzal járt, hogy az iskolát ideiglenesen bezárták.

- Egy 19 éves ausztrál férfi, William Yeates bűnösnek vallotta magát, mert 2024 és 2025 között MI segítségével készített és terjesztett szexuálisan explicit DeepFake képeket egy áldozatról. Az ügy érdekessége, hogy ez az első olyan büntetőeljárás, amit az új, MI-alapú visszaélések ellen hozott törvények alapján bírálnak el Ausztráliában. A jogszabályok értelmében a maximális büntetés akár 7 év börtön is lehet.

Szélsőséges esetek

- Egy idős férfit kerestek meg csalók DeepFake képekkel és videókkal, ráadásul a technológia segítségével képesek voltak a hatóságnak kiadni magukat. A támadók célja a pszichológiai nyomásgyakorlással a pénzszerzés volt. Az állandó fenyegetés miatt folyamatos stresszben és félelemben élt, ami öngyilkos közeli állapotba juttatta a férfit, de a rendőrségi beavatkozásnak hála sikerült megakadályozni a tragédiát.

- Egy másik esetben egy Elijah Heacock nevű 16 éves fiú követett el öngyilkosságot, miután egy ismeretlen személy megzsarolta őt egy MI által generált meztelen fotóval, ami őt ábrázolta. A zsaroló 3000 dollárt követelt, különben azzal fenyegette a fiút, hogy elküldi a képet a barátainak és a családjának. Elijah a fenyegetéstől pánikba esett és nem sokkal később öngyilkos lett, szülei azóta tudatosító előadásokat tartanak a témában.

A CBS NEWS cikke szerint ez a jelenség egyáltalán nem ritka, egy év alatt közel 500.000 ehhez hasonló esetet jelentenek kiskorúak vonatkozásában, az FBI adatai szerint pedig 2021 óta eddig közel 20 fiatal halt meg ilyen és ehhez hasonló zsarolások miatt.

Létezik egyáltalán DeepFake elleni védelem?

A védelmi megoldások erősen függenek a felhasználók aktuális szerepétől, ugyanis nem mindegy, hogy tartalomfogyasztó felhasználóként igyekszik valaki eldönteni, hogy az általa látottak valóban megtörténtek-e vagy a szórakozás kedvéért készít valaki a barátjáról egy vicces videót, vagy netán az áldozat igyekszik megakadályozni, hogy tovább terjedjen egy róla készült DeepFake tartalom.

- Tartalomfogyasztás szempontjából – amennyiben nincs jelölve egy adott médiáról, hogy az DeepFake – érdemes figyelmet fordítani a részletekre. Egyrészt nézzük vagy hallgassuk meg minél többször és alaposan az adott tartalmat, furcsa mimikák, fények, hangok és egyéb torzulások után kutatva, másrészt pedig keressünk más forrásokat a témát illetően.

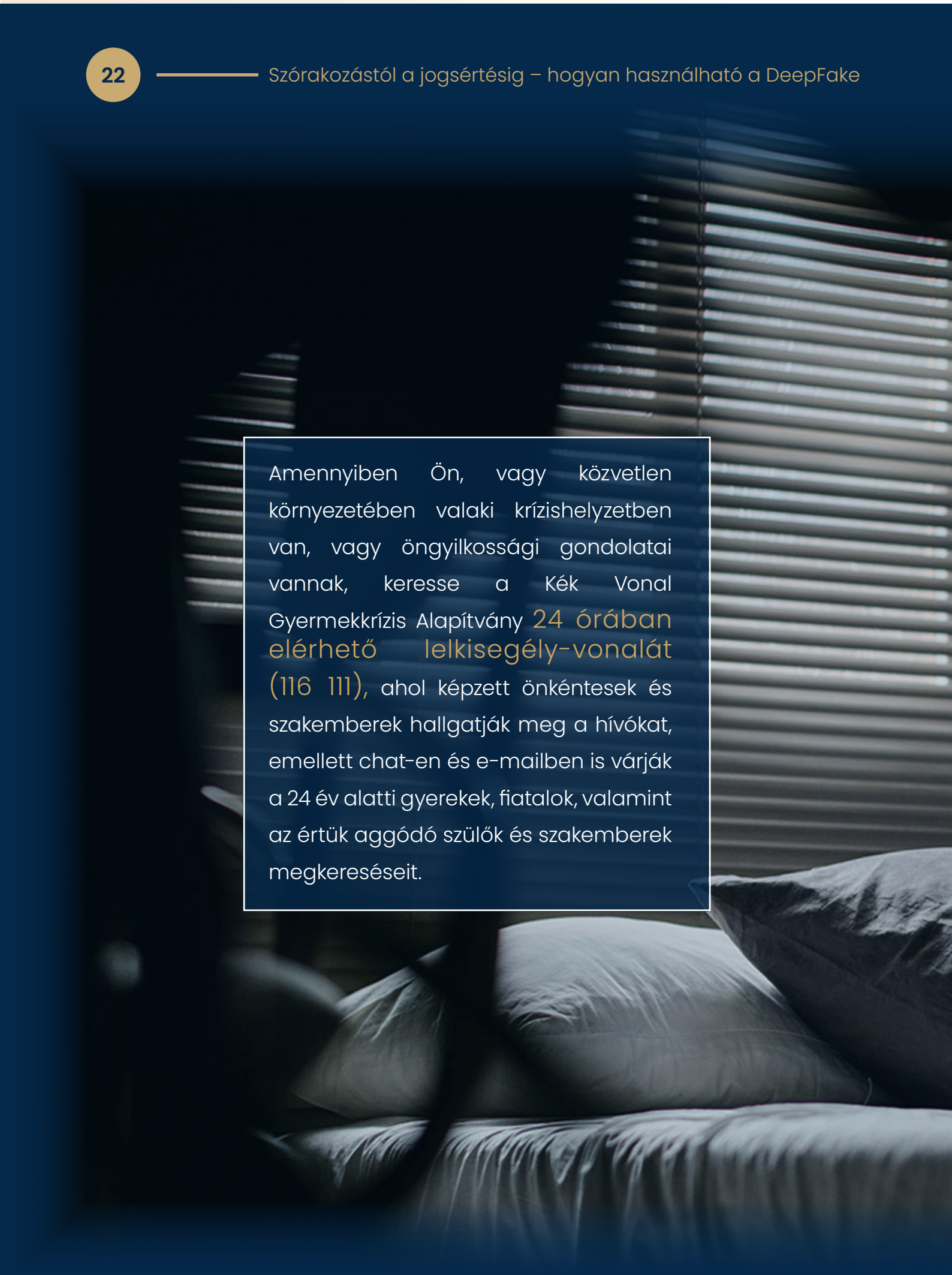
  Amennyiben valaki nem bízik a saját ítélő képességében, használhat különböző DeepFake detektáló eszközöket. Ilyen például a Sensity AI, Deepguard.ai és a McAfee Deepfake Detector, de számos más online és offline, ingyenes és fizetős verzió érhető el a piacon.

  Ha egy adott tartalomról a fenti segítségekkel sem tudjuk megállapítani, hogy az valóban megtörtént-e vagy csak egy DeepFake, úgy érdemes fenntartásokkal kezelni az általa látottakat és hallottakat. Ehhez nyújthat segítséget az Intézetünk és a Sans Institute által közösen kiadott OUCH! egy korábbi, Sajátítsunk el egy új túlélési készséget: ismerjük fel a deepfake-eket! című kiadványa, amelyben ismertetésre kerülnek azok az apró kis részletek, amelyek segíthetnek leleplezni a DeepFake tartalmakat.

- Tartalomkészítés szempontjából, a lehető legfontosabb az átláthatóság, vagyis, hogy mindig, minden körülmény között fel kell tüntetni, hogy az adott tartalmat mesterséges intelligencia hozta létre. DeepFake tartalom létrehozásakor érdemes arra ügyelni, hogy az elkészített tartalom ne legyen egyéb módon se jogsértő, például valakinek a megszemélyesítésével készült kép, videó vagy hang készítéséhez mindig kérjük előzetesen az illető beleegyezését. Ne készítsünk ártó szándékú, rossz hírnévkeltő vagy félrevezető tartalmakat!

- Az áldozatok számára egyre több eszköz áll rendelkezésre a károk minimalizálása érdekében, ami részben az ehhez szükséges szabályozási kereteknek és a már kialakult folyamatoknak köszönhető. A többlépcsős reakció első lépése a bizonyítékgyűjtés (ki, mit, milyen platformon tett közzé), majd jelenteni a tartalmat a platform üzemeltetőinek. Ezt szerencsére a legtöbb platform egyre egyszerűbbé teszi. Ezt követően jöhetnek a megfelelő jogi lépések, mint a rendőrségi feljelentés vagy – a jogsértés függvényében – az illetékes szervezeteknek való bejelentés. Ilyen lehet például a személyes adatok védelmével visszaélő tartalmak, amelyeket a Nemzeti Adatvédelmi és Információszabadság Hatóságnak (NAIH) kell bejelenteni. Végül érdemes nyilvánosan is tisztázni a megtörténteket a reputáció védelme érdekében.

• Végül a megelőzésről is érdemes néhány gondolatot megejteni, bár igazából nincs jól bevált recept arra vonatkozóan, hogy ne állítsanak elő DeepFake tartalmat rólunk. Mégis csökkenthető a kockázat, ha minél kevesebb biometrikus (kép, videó, hang) adatot teszünk közzé magunkról. Javasolt rendszeres időközönként ellenőrizni, hogy milyen információk érhetők el rólunk nyilvánosan és az általunk nem jóváhagyott tartalmakat jelenteni az adott platformokon és kérni azok mielőbbi eltávolítását.



Amennyiben Ön, vagy közvetlen környezetében valaki krízishelyzetben van, vagy öngyilkossági gondolatai vannak, keresse a Kék Vonal Gyermekkrízis Alapítvány 24 órában elérhető lelkisegély-vonalát (116 111), ahol képzett önkéntesek és szakemberek hallgatják meg a hívókat, emellett chat-en és e-mailben is várják a 24 év alatti gyerekek, fiatalok, valamint az értük aggódó szülők és szakemberek megkereséseit.

Források

NBSZ NKI *Mekkora fenyegetést jelentenek valójában a deepfake-ek?* (2022) https://nki.gov.hu/it-biztonsag/elemzesek/mekkora-fenyegetest-jelentenek-valojaban-a-deepfake-ek/ (Letöltve: 2026.04.08.)

UNIVERSITY OF VIRGINIA *What the heck is a deepfake?* https://security.virginia.edu/deepfakes (Letöltve: 2026.04.08.)

IBM *Weaponizing reality: The evolution of deepfake technology* https://www.ibm.com/think/x-force/weaponizing-reality-evolution-deepfake-technology (Letöltve: 2026.04.08.)

National Geographic *These manipulated photos are the original political deepfakes* (2024) https://www.nationalgeographic.com/history/article/political-photo-manipulation-in-history (Letöltve: 2026.04.29.)

SECURITY HERO *2023 STATE OF DEEPFAKES* https://www.securityhero.io/state-of-deepfakes/ (Letöltve: 2026.04.08.)

GRAPHITE *More Articles Are Now Created by AI Than Humans* https://graphite.io/five-percent/more-articles-are-now-created-by-ai-than-humans (Letöltve: 2026.04.08.)

Ahrefs blog *74% of New Webpages Include AI Content* (2025) https://ahrefs.com/blog/what-percentage-of-new-content-is-ai-generated/ (Letöltve: 2026.04.29.)

JOGÁSZVILÁG *Mesterséges intelligencia és bizonyítási kérdések a perben* (2025) https://jogaszvilag.hu/szakma/mesterseges-intelligencia-es-bizonyitasi-kerdesek-a-perben/ (Letöltve: 2026.04.09.)

EURÓPAI UNIÓ *AZ EURÓPAI PARLAMENT ÉS A TANÁCS (EU) 2024/1689 RENDELETE 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) (2024) https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=OJ:L_202401689 (Letöltve: 2026.04.08.)

NEMZETI JOGSZABÁLYTÁR *2025. évi LXXV. törvény az Európai Unió mesterséges intelligenciáról szóló rendeletének magyarországi végrehajtásáról* (2025) https://njt.hu/jogszabaly/2025-75-00-00 (Letöltve: 2026.04.09.)

NEMZETI JOGSZABÁLYTÁR *2012. évi CCXXIII. törvény a Büntető Törvénykönyvről szóló 2012. évi C. törvény hatálybalépéséhez kapcsolódó átmeneti rendelkezésekről és egyes törvények módosításáról* (2023) https://njt.hu/jogszabaly/2012-223-00-00 (Letöltve: 2026.04.09.)

TÖRÖK BALÁZS *A generatív AI lehetséges hatásai az online hirdetésekben* (2024) https://www.torokbalazs.com/blog/a-generativ-ai-lehetseges-hatasai-az-online-hirdetsekben (Letöltve: 2026.04.09.)

THEAIPAGE INSTAGRAM *Historical Icons Brought Back to Life using Artifical Intelligence* (2025) https://www.instagram.com/reel/DH1ZObGSmNv/ (Letöltve: 2026.04.09.)

ICTGLOBAL *MI-influenszerek: a virtuális idolok forradalma* (2023) https://ictglobal.hu/iparagi-megoldasok/mi-influenszerek-a-virtualis-idolok-forradalma/ (Letöltve: 2026.04.09.)

INSTAGRAM https://www.instagram.com/lilmiquela/?hl=hu (Letöltve: 2026.04.10.)

INSTAGRAM https://www.instagram.com/miazelu/ (Letöltve 2026.04.10.)

NATIONAL LIBRARY OF MEDICINE *Promising for patients or deeply disturbing? The ethical and legal*

aspects of deepfake therapy (2024) https://pmc.ncbi.nlm.nih.gov/articles/PMC12322411/ (Letöltve: 2026.04.10.)

NBSZ NKI *AI által támogatott social engineering: új korszak a kibertámadások világában* (2026) https://nki.gov.hu/it-biztonsag/hirek/ai-altal-tamogatott-social-engineering-uj-korszak-a-kibertamadasok-vilagaban/ (Letöltve: 2026.04.10.)

DEEPFAKE.AI *What Legal and Ethical Challenges Does Deepfake Technology Pose to Society?* (2026) https://www.deep-fake.ai/en/blog/deepfake-ethics-law (Letöltve: 2026.04.10.)

CORVINÁK *Hogyan illeszkednek a deepfake tartalmak a post-truth politika fogalmába?* (2025) https://corvinak.hu/cikk/2025/04/16/hogyan-illeszkednek-a-deepfake-tartalmak-a-post-truth-politika-fogalmaba (Letöltve: 2026.04.13.)

NBSZ NKI *KiberKedd: a BEC-típusú online csalások – maradjunk elővigyázatosak a munkahelyünkön is!* (2025) https://nki.gov.hu/it-biztonsag/tanacsok/kiberkedd-a-bec-tipusu-online-csalasok-maradjunk-elovigyazatosak-a-munkahelyunkon-is/ (Letöltve: 2026.04.16.)

EURONEWS *Hamis videón szólítja fel megadásra katonáit az ukrán elnök* (2022) https://hu.euronews.com/2022/03/17/hamis-videon-szolitja-fel-megadasra-katonait-az-ukran-elnok (Letöltve: 2026.04.16.)

ITBUSINESS *Lecsaptak az X-re Taylor Swift rajongóinak százezrei* (2024) https://itbusiness.hu/technology/aktualis-lapszam/ict-market/taylor-swift-deepfake/ (Letöltve: 2026.04.16.)

THEGUARDIAN *Taylor Swift AI images prompt US bill to tackle nonconsensual, sexual deepfakes* (2024) https://www.theguardian.com/technology/2024/jan/30/taylor-swift-ai-deepfake-nonconsensual-sexual-images-bill (Letöltve: 2026.04.16.)

ARS TECHNIKA *Explicit deepfake scandal shuts down Pennsylvania school* (2024) https://arstechnica.

com/tech-policy/2024/11/school-failed-to-report-ai-nudes-of-kids-for-months-now-parents-are-suing/ (Letöltve: 2026.04.16.)

NEWS.COM.AU *Private school graduate admits deepfake offence* (2026) https://www.news.com.au/national/south-australia/private-school-graduate-admits-deepfake-offence/news-story/3e5dcca948f08c5c36d8a608277d652e (Letöltve: 2026.04.16.)

CBS NEWS *A teen died after being blackmailed with A.I.-generated nudes. His family is fighting for change* (2025) https://www.cbsnews.com/news/sextortion-generative-ai-scam-elijah-heacock-take-it-down-act/ (Letöltve: 2026.04.16.)

OECD.AI *Deepfake Extortion Pushes Elderly Man to Brink of Suicide in Ghaziabad* (2023) https://oecd.ai/en/incidents/2023-11-30-bb66 (Letöltve: 2026.04.16.)

TELEX *A német televíziózás álompárjának tűntek, aztán kiderült a férj sötét titka* (2026) https://telex.hu/kulfold/2026/03/29/nemetorszag-rendorseg-nyomozas-deepfake-porno-collien-fernandes-christian-ulmen-digitalis-eroszak (Letöltve: 2026.04.22.)

SENSITY AI *Forensic Deepfake Detection* https://sensity.ai/ (Letöltve: 2026.04.22.)

Deepguard.ai *Unwavering Security and Peace of Mind* https://www.deepguard.ai/ (Letöltve: 2026.04.22.)

McAfee *McAfee® Deepfake Detector flags AI-generated audio within seconds* https://www.mcafee.com/ai/deepfake-detector/ (Letöltve: 2026.04.22.)

Kék Vonal *Keress minket bátran!* https://kek-vonal.hu/ (Letöltve: 2026.04.22.)

FBI *Sextortion: A Growing Threat Targeting Minors* (2024) https://www.fbi.gov/contact-us/field-offices/nashville/news/sextortion-a-growing-threat-targeting-minors (Letöltve: 2026.04.29.)

YouTube *Deepfake video of Volodymyr Zelensky surrendering surfaces on social media* (2022) https://www.youtube.com/watch?v=X17yrEV5sl4 (Letöltve:2026.05.07.)

NBSZ NKI *Sajátítsunk el egy új túlélési készséget: ismerjük fel a deepfake-eket!* (2022) https://nki.gov.hu/wp-content/uploads/2022/03/ouch_march_2022_hungarian_learn_a_new_survival_skill_spotting_deepfake.pdf (Letöltve: 2026.05.08.)

NEMZETI
KIBERBIZTONSÁGI
INTÉZET

Kibertámadás!
podcast

Nemzetbiztonsági Szakszolgálat
Nemzeti Kiberbiztonsági Intézet

titkarsag@nki.gov.hu

nki.gov.hu

+36 (1) 325 7672

2026