G. Karácsony Gergely

„Ideje megváltoztatni” – Az arcfelismerő rendszerek alkalmazásának alapjogi kockázatai a közösségi média és a deepfake korában

A technológiai fejlődés minden korábbinál pontosabb távoli arcfelismerő rendszerek megalkotását tette lehetővé. Ezekkel a rendszerekkel nap mint nap találkozunk: az okostelefonunk így ismer fel minket, így lépünk be a munkahelyünkre, és az utcai térfigyelő kamerarendszer mögött is egy ilyen algoritmus áll. Ebben a tanulmányban arra hívjuk fel a figyelmet, hogy – noha népszerűségének csúcsát éli éppen – az arcfelismerő rendszer használata számos jogi kockázatot rejt magában, ráadásul a világhálón bárkiről rendelkezésre álló nagy mennyiségű képanyag felhasználásával olyan pontos hamisítványok készíthetők egy személy kinézetéről, hogy a fejlett arcfelismerő rendszereket is megtévesztik. Ezért az arcfelismerő technológia működésének szabályozására és az arcmással való visszaélés elleni hatékony küzdelmet segítő megoldásokra teszünk javaslatot ebben a tanulmányban.

Bevezetés

Az emberi arc az egyik legpontosabb és leghatékonyabb azonosítójegyünk. Azt mondják, hogy nincs két egyforma arc a világon, de vajon elgondolkodtak már azon, hogy miért alakult ki nálunk ez a nagy fokú változatosság, amíg más állatfajok egyedeinél a fej formája és színezete csak kismértékű eltéréseket mutat? Kutatások bizonyítják, hogy az emberi arc az evolúció során úgy fejlődött, hogy kifejezetten nagy variabilitást mutasson az egyes arcrészek mérete, színe, alakja és egymáshoz viszonyított helyzete. Ennek pedig éppen az azonosítás miatt van szerepe fajtársaink körében.[1] Ahogy az emberek számára az arc az egyik leginkább megbízható, távolról is leolvasható azonosító jellegzetesség, ezzel ellentétben az informatikai rendszerek esetében egyre inkább úgy tűnik, hogy minden megoldás jobb, mint az arcfelismerés. Ebben a tanulmányban összefoglaljuk azokat a jellemzőket, amelyek az arcfelismerő megoldások működését technológiai szempontból meghatározzák, illetve bemutatjuk a technológia jelenlegi alkalmazási területeit, kitérve annak kockázataira és korlátaira. A technológia működésével kapcsolatos műszaki információk ismertetését elengedhetetlenül szükségesnek tartom annak érdekében, hogy a jogi

[1] Michael J. Sheehan – Michael W. Nachman: Morphological and Population Genomic Evidence that Human Faces have Evolved to Signal Individual Identity. *Nature Communications,* (2014), 5.

https://doi.org/10.36250/01065_16

problémákat és a szabályozási kérdéseket meg tudjuk érteni. A tanulmányban ráirányítjuk a figyelmet arra a problémára, amelyet az arcfelismeréssel történő azonosítás elterjedése és ezzel egyidejűleg az úgynevezett *deepfake*[2]-technológia tökéletesedése együttesen jelent a közösségi hálózatok által dominált digitális térben. A tanulmány az arcfelismerő technológiák kockázatai kapcsán a kérdéskör emberi jogi aspektusát felismerő és ezt a nyilvános demokratikus diskurzus részeként kezelő jogállamok keretei között vizsgálódik, ezért nem tér ki részletesen az úgynevezett „megfigyelő állam" problémakörére, amely egészen más körülmények között, az emberi jogok jelentős korlátozásával valósul meg.

Azonosítás, hitelesítés és az arckép szerepe

Az információs rendszerek működésében szinte az első számítógépek megjelenése óta jelen lévő probléma az azonosítás kérdése. A többfelhasználós operációs rendszerek felépítéséből következő alapvető biztonsági követelmény volt az, hogy a különböző jogosultsági szinttel rendelkező felhasználókat megkülönböztessék egymástól, illetve az adott rendszerbe (számítógépre) belépési jogosultsággal rendelkezőket elválasszák a belépésre nem jogosultaktól. Ennek a legegyszerűbb – és máig leginkább elterjedt – módja az, hogy a felhasználóknak saját, egyedi belépési *(login)* nevet kell megadniuk, és a jogosultságok ehhez a névhez kapcsolódnak. Az azonosítás *(identification)* folyamata során tehát a bejelentkezéssel próbálkozó személy tesz egy állítást arról, hogy ő kicsoda (például beírja a felhasználónevét vagy a regisztrációnál megadott e-mail-címet).[3] Ha jogosult a belépésre, illetve az adott erőforráshoz való hozzáférésre, akkor a rendszer beengedi; ha pedig nem jogosult, akkor megtagadja a hozzáférést. Ez a lépés önmagában azonban nem garantálja a biztonságot, hiszen egy rosszindulatú személy is állíthatja magáról, hogy ő egy jogosult felhasználó, akkor is, ha nem az.

Már a Bibliában is találunk említést arról, hogy jelszót használtak az idegenek kiszűrésére és a közéjük tartozók azonosítására: „[…] akkor felszólították: Mondd ki tehát: »Sibbólet« – ami annyit jelent, mint kalász. Aki azt felelte: »Szibbólet«, mert ugyanazt a hangot a kalász szóban kiejteni nem tudta, azt legott megfogták s ott a Jordán átkelőjénél megölték."[4] Az informatikában hitelesítésnek *(authentication)*[5]

[2] A kifejezés a mélytanulás *(deep learning)* és a hamisítvány *(fake)* szavak kombinációjából ered, és azt a módszert takarja, hogy a mesterséges intelligencia segítségével egy személyről rendelkezésre álló képek vagy hangfelvételek alapján készítenek manipulált felvételt. Lásd bővebben a *Deepfake – a mesterséges intelligencia használata az arcfelismerés megtévesztésére* fejezet alatt.

[3] Matematikai szimbólumokkal ezt nevezzük 1:N azonosításnak, vagyis amikor egy személyt választunk ki egy nagy halmazból.

[4] Bírák könyve, 12:6.

[5] Szimbólumokkal leírva ez az 1:1 azonosítás, vagyis amikor egy személyről kapott azonosító információt feleltetjük meg a róla korábban eltárolt adatokkal.

nevezik azt a lépést, amikor meggyőződünk arról, hogy a bejelentkezni próbáló személy valóban az, akinek az azonosítás során mondta magát. Erre többféle megoldást kidolgoztak már, amelyek lényege az, hogy az azonosítás során megadott adatot összekötik egy olyan információval, amellyel feltehetőleg csak a megjelölt személy rendelkezik. A hitelesítő tényező *(authentication factor)* lehet olyasvalami, amit az érintett személy tud (például jelszó), valami, amivel ő rendelkezik (például mágneskártya, okostelefon, USB-kulcs), vagy pedig valamilyen személyes jellegzetessége (ujjlenyomat, arckép, hang stb.). A jelszóval való azonosítás viszonylag egyszerű, és nem követel túl nagy erőfeszítést az azonosítani kívánt személytől: csak emlékeznie kell a megadott titkos kódra. Az egyszerűsége azonban a veszélye is: a legtöbb jelszó könnyen kitalálható,[6] illetve megfelelő számítókapacitás használata mellett hamar feltörhető.[7] A bonyolultabb jelszavak pedig magukban hordozzák annak veszélyét, hogy a felhasználó elfelejti őket, vagy pedig felírva tárolja valahol, amelyet könnyen megszerezhet egy rosszindulatú fél. A birtokunkban lévő eszközök belépőként való használatára számos példát tudunk bemutatni a mágneskártyától a mobiltelefonra érkező, egyszeri kódot tartalmazó szöveges üzenetig, de gondolhatunk a lakáskulcsunkra is. Ezt a módszert távolról aligha lehet feltörni, azonban a hitelesítő eszköz elvesztése, ellopása, illetve lemásolása esetén a támadó könnyűszerrel kiadhatja magát a jogosult felhasználónak, és beléphet a védett rendszerbe.

A hitelesítő faktorok közül a jogosult személyes jellegzetességeit alapul vevő megoldások tűnnek a leginkább biztonságosnak, hiszen az ujjlenyomat, arckép vagy hangminta használatához a tényleges személy valódi jelenléte szükséges, ezek a jellemzők nem lophatók el, és nem is törhetők fel úgy, mint a jelszavak. Éppen ezért egyre több helyen használják az informatikában, illetve védett objektumokba történő beléptetés során a biometrikus azonosítást.

Az arcalapú azonosítás, jelentőségét tekintve, kiemelkedik az egyén testi jellemzőit alapul vevő többi hitelesítési megoldás közül. Ennek oka a relatív egyszerűsége és az, hogy érintésmentes, távolról történő azonosítást is lehetővé tesz. Ahogyan azt a bevezetőben is írtuk, az emberi arc egy kifejezetten azonosítási célra kialakult evolúciós jellegzetességünk, aminek az előnye, hogy viszonylag nagy messzeségből is „működik", ránézésre is felismerjük az ismerős arcokat a tömegben. Ez a hatás pedig elvezet minket a távoli arcfelismerés mesterséges intelligencia által megerősített és felgyorsított világába. Az elmúlt néhány év sorra hozta a távoli arcfelismerő rendszerekkel kapcsolatos technológiai fejlesztésekről, hasznos vagy kockázatos

[6] A leggyakrabban használt jelszavak 2021-ben, mintegy 15 milliárd, ismertté vált jelszó elemzése alapján: 123456, 123456789, qwerty. Emellett gyakoriak a születési dátumok, keresztnevek és káromkodások is a jelszavak között. Forrás: Bernard Meyer: Most Common Passwords: Latest 2022 Statistics. *Cybernews,* 2022. május 27.

[7] E cikk megírása idején egy 8 karakteres, számokat, valamint kis- és nagybetűket tartalmazó jelszó feltöréséhez nagyjából egy órára van szükség. Forrás: Marina Watts: Read This Chart to Find Out How Fast Someone Can Hack Into All of Your Accounts. *Newsweek,* 2020. január 9.

alkalmazásukról szóló híreket.[8] Ma már nem okoz meglepetést az, hogy a repülőtereken a biztonsági ellenőrzés része az arcfelismerő kapun való áthaladás, amely validálja az igazolvány útján bemutatott személyazonosságot. Ezek a rendszerek az Egyesült Államokban a 2000-es évek elejétől,[9] az EU országaiban pedig a 2010-es évektől vannak jelen.[10] E tanulmány megírásakor épp aktuális hír, hogy a moszkvai metró beléptetőkapuit arcalapú menetdíjfizetési rendszerrel látták el, amely képes felismerni a belépő utast, és a rendszerben regisztrált bankkártyáról automatikusan levonni a viteldíjat, így fizikailag létező jegyre vagy bérletre, illetve pénzbedobásra sincs tovább szükség.[11] Nem maradtak el a fejlesztésben az online platformok sem: a legismertebb közösségi oldalak és technológiai szolgáltatók egyaránt kifejlesztették és működtetik az arcfelismerő mechanizmusukat a rendszerbe feltöltött képeken és videókon.[12]

Az arcalapú távoli azonosítás jelentősége hatványozottan jelentkezik a városi térfigyelő kamerarendszerek folyamatos kiépülésével és terjedésével. A mesterséges intelligencián alapuló arcfelismerő rendszerek a térfigyelő kamerákkal kombinálva soha nem látott módon hatékony eszközt adtak a bűnüldöző szervek kezébe. Az ilyen rendszerek alkalmazásával hamar azonosítani képesek egy bűncselekmény elkövetőjét, illetve nyomon követhetik a menekülő bűnözők útját. Az eltűnt személyek felkutatására, illetve az emberrablás és emberkereskedelem áldozatainak megsegítésére szakosodott szervezetek szintén üdvözlik az arcfelismeréssel ellátott kamerarendszereket, amelyek sokszor segítenek régóta eltűnt vagy bűncselekmény áldozatává vált emberek azonosításában és megtalálásában. A magánszféra védelméért aggódók viszont nem nézik jó szemmel ezt a technológiát, amely a személyek folyamatos megfigyelését és korlátlan adatgyűjtést is lehetővé teszi, visszaélésszerűen használva pedig megvalósítja a „megfigyelő állam" *(surveillance state)* disztópiáját.

Az iPhone X és a vele együtt debütáló Face ID megoldás[13] 2017-es megjelenése óta a szórakoztató elektronika és a mobiltechnológia területén is folyamatosan terjed a felhasználó arcával történő azonosítás az eszközhöz való hozzáférés érdekében. A bűnüldöző szervek és hatóságok, valamint a közlekedési cégekhez hasonló nagyvállalatok mellett az arcalapú, illetve más biometrikus jellemzőket felhasználó azonosítást számos kisebb szervezet is alkalmazza. Egyetemi kollégiumokba, sportlétesítményekbe, edzőtermekre és munkahelyekre is sokszor már ujjlenyomattal vagy

[8] Lásd például: Neil Hughes: A Backlash Against Facial Recognition is Brewing. *TechHQ,* 2019. augusztus 20.

[9] Francesca Street: How Facial Recognition is Taking over Airports. *CNN Travel,* 2019. október 8.

[10] Jose Sanchez del Rio – Daniela Moctezuma – Cristina Conde – Isaac Martin de Diego – Enrique Cabello: Automated Border Control E-Gates and Facial Recognition Systems. *Computers & Security,* 62. (2016). 49–72.

[11] Lásd Pjotr Sauer: Privacy Fears as Moscow Metro Rolls out Facial Recognition Pay System. *The Guardian,* 2021. október 15.

[12] Lásd például: Mi az arcfelismerési beállítás a Facebookon és hogyan működik? Facebook.

[13] Lásd Apple: A Face ID használata iPhone és iPad Pro készüléken.

arcfelismeréssel lehet belépni, a legtöbbször egyelőre csak kiegészítő lehetőségként. Ahogy a technológia egyre fejlettebb lesz, és ezzel együtt egyre olcsóbban elérhető, hamar begyűrűzik a hétköznapok használati tárgyai közé: az interneten alig több mint tízdolláros áron rendelhető olyan otthoni biztonsági kamera, amely felismeri a családtagokat; egyes okostévék pedig arc alapján azonosítják az előttük ülő nézőt, hogy személyesen neki ajánlhassanak tartalmakat.[14]

A téves azonosítás problematikája

A biometrikus azonosító rendszerek – és így az arcfelismerő megoldások is – két irányban képesek tévedni. A hibás negatív eredmény *(type 1 error)* azt jelenti, hogy a rendszer nem ismer fel valakit, aki egyébként szerepel az adatbázisban. Ezzel szemben a hibás pozitív eredmény *(type 2 error)* esetén a rendszer pozitív eredményt jelez olyasvalaki esetében, aki valójában nem szerepel az adatbázisban. Ha az arcfelismerő rendszer beléptetéshez kapcsolódik, akkor a hibás negatív eredmény okoz általában a felhasználónak kellemetlenséget, hiszen ilyen esetben újra meg kell próbálkoznia az azonosítással, másik hitelesítési módot kell választania, vagy pedig emberi közreműködést kell kérnie. Ritkán fordul elő itt súlyos következmény, esetleg az képzelhető el, hogy többszöri hibás próbálkozással valaki tartósan kizárja magát a rendszerből. A hibás pozitív eredmény ezzel szemben a védett rendszerre nézve jelent kockázatot, hiszen ez esetben a beléptető algoritmus olyasvalakit enged be, akinek nem volt jogosultsága belépni, és akár rosszindulatú aktor is lehet. Ez a hiba nemcsak nehezen javítható, hanem nehezen is ismerhető fel, hiszen itt nincs, aki felhívja a rendszer kezelőjének a figyelmét a téves döntésre, az algoritmus pedig nem végez önellenőrzést. Ezért a beléptetésre használt biometrikus azonosító rendszerek esetében általában az az elvárás, hogy a hibás pozitív eredmények előfordulását *(FPR – false positive rate)* szorítsák a minimálisra, még annak az árán is, hogy így esetleg növekszik a hibás negatív döntések aránya, amellyel szemben ebből fakadóan nagyobb toleranciát lehet megengedni.[15]

Egy másik élethelyzetben megvizsgálva a hibalehetőségeket, egészen más kép tárul a szemünk elé. Ha az arcfelismerést személyek keresésére vagy bűnelkövetők azonosítására, illetve hasonló súlyos következményekkel járó célra használjuk, akkor mást mutat az a kérdés, hogy kit sújt jobban a téves eredmény. A hibás negatív eredmény itt azt jelenti, hogy a rendszer nem ismert fel és nem szűrt ki egy potenciális

[14] Samsung: *What are the Requirements for Using Face Recognition in Samsung H Series TV(UA40H6400)?* 2020. október 14.

[15] Így például az európai és nemzetközi légügyi hatóságok szabványai a repülőtéri arcfelismerő kapuk esetében 5%-os hibás negatív döntést engednek meg, míg a hibás pozitív döntések aránya nem haladhatja meg a 0,1%-ot. Lásd például: Frontex: *Best Practice Technical Guidelines for Automated Border Control (ABC) Systems.* 2015.

rosszindulatú személyt vagy bűnözőt, ami a rendszer által védett értékekre jelent kockázatot. A hibás pozitív eredmény esetén viszont a rendszer ismert bűnözőként azonosít egy ártatlan embert, vagy pedig egy ismeretlen elkövetőt köt össze egy harmadik személlyel, akire esetleg hasonlít. Ez utóbbi esetekben mindenképp a tévedésből megnevezett személy viseli az esetleges negatív következményeket, akár büntetőeljárásban is védekezésre kényszerülve. Jól látható, hogy ezúttal az egyéni jogok és szabadságok védelme szempontjából fűződik ahhoz erősebb érdek, hogy a hibás pozitív eredményeket minimálisra csökkentsük.

A hibás pozitív azonosítási eredmény jogi kockázatai

A fent bemutatott lehetséges felismerési hibák közül a legfontosabb a hibás pozitív eredmény következményeit és lehetséges hatásait megvizsgálni, hiszen az egyének alapjogai és szabadsága szempontjából ez a fajta hiba hordozza a legnagyobb kockázatot. Az alábbiakban bemutatjuk a hibás pozitív azonosítás kockázatait a technológia különböző alkalmazási területein. Kitérünk a hibás azonosítást elősegítő, az arc- és hangmintát meghamisító fejlett technológia, az úgynevezett deepfake működésére is.

Arcfelismerési rendszerek a bűnüldözésben

A bűnüldöző szervek mára egyre több esetben használnak arcfelismerő rendszereket a bűncselekmények elkövetőinek azonosítására vagy a szökésben lévő bűnözők nyomon követése, illetve elfogása céljából. Az Egyesült Államokban kiterjedt vizsgálat alá vették a leggyakrabban használt arcfelismerő rendszereket, és arra jutottak, hogy sok közülük viszonylag magas téves azonosítási aránnyal dolgozik, különösen a kisebbségekhez tartozó személyek azonosítása esetében.[16] A téves azonosítás kapcsán ebben az esetben a leggyakrabban felmerülő kérdés az, hogy mit tehet az egyén, akit az arcfelismerő rendszer tévesen azonosított, és emiatt egy bűncselekmény gyanúsítottjává vált. Az elmúlt néhány évben számos ilyen ügy vált ismertté, többek között Nijeer Parks esete az Egyesült Államokban, aki 10 napot töltött börtönben, és mintegy 5000 dolláros ügyvédi költsége merült fel amiatt, mert egy arcfelismerő rendszer egy bolti lopás és a menekülés közbeni, rendőrök elleni támadás elkövetőjeként azonosította.[17] A téves azonosítás elsőként ismertté vált áldozata, Robert Williams 30 órát töltött a rendőrségi fogdán, majd pedig

[16] Lásd: Patrick Grother et al.: *Face Recognition Vendor Test (Frvt) Part 3: Demographic Effects.* National Institute of Standards and Technology, U.S. Department of Commerce, Nistir 8280, 2019. december.

[17] Kashmir Hill: Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match. *The New York Times,* 2020. december 29.

hosszas pereskedés során sikerült tisztáznia magát, annak ellenére, hogy alibivel is rendelkezett a terhére rótt bűncselekmény elkövetésének idejére.[18] A bűnüldöző szervek ezekben az esetekben azt hangsúlyozzák,[19] hogy az arcfelismerő rendszerek önmagukban nem adnak bűnösséget megalapozó bizonyítékot, pusztán nyomot *(investigative lead)* szolgáltatnak a hatóságok számára, akik más bizonyítékokat is felhasználva döntenek a gyanúsításról. A legtöbb rendszer valóban egy százalékos bizonyossági fokot jelöl meg, hogy a feltöltött képen mekkora valószínűséggel szerepel a keresés eredményeként kiválasztott személy. A kockázatot leginkább azok az esetek hordozzák magukban, amikor nem áll rendelkezésre más, meggyőző bizonyíték, vagy ha az azonosítás közvetett módon történik.[20]

A téves azonosítás alapján elinduló folyamat egy olyan spirálba kényszeríti bele a szereplőket, amelyből csak azon az úton lehet kitörni, ha a folyamatot elindító információt, vagyis a távoli biometrikus azonosítás eredményét kérdőjelezzük meg. Erre gyakran a nyomozati szakasz során nincsen megfelelő lehetőség, sokszor csak a bírósági szakaszban mérlegelik újra a bizonyítékokat. A rendvédelmi szervek által használt arcfelismerő rendszerek különféle fejlesztő és gyártó cégektől származnak, így eltérő működési elvük és pontosságuk van, a belső mechanizmusukat pedig gyakran üzleti titokként kezelik. Ugyanígy bizonytalanná teszi a rendszerek működését az, hogy az azonosításhoz használt képek minden rendszerben más és más forrásból származnak, ezért a megbízhatóságuk is jelentősen eltér.[21] Az Európai Unió mesterséges intelligenciával kapcsolatos rendelettervezete[22] a természetes személyek távoli biometrikus azonosítására szolgáló MI-rendszereket éppen azok technikai pontatlansága és az így adott torzított eredmények miatt nagy kockázatúnak minősíti, és előírja, hogy azokra a naplózási képességekkel és az emberi felügyelettel kapcsolatos különös követelményeknek kell vonatkozniuk.[23] Így kimondja

[18] Kashmir Hill: Wrongfully Accused by an Algorithm. *The New York Times,* 2020. június 24.

[19] Lásd például: Facial Recognition – Frequently Asked Questions. *State of Michigan,* 2019. szeptember.

[20] Az ismertetett esetben a rendőrség egy olyan személyt idézett szemtanúként, aki nem volt ott a bűncselekmény helyszínén, hanem maga is csak a videófelvételt látta, és ez alapján nyilatkozott arról, hogy az elkövető hasonlít a meggyanúsított személyre. Később a belső vizsgálatban ezt eljárási szabálytalanságnak ítélték.

[21] A közelmúltban kavart nagy botrányt a számos rendvédelmi szervnek arcfelismerő rendszert szolgáltató Clearview AI nevű cég esete, amely a felismeréshez használt képeket a Facebook, YouTube és más platformok – azok tudta és a felhasználók beleegyezése nélküli – átfésülése és lemásolása *(scraping)* útján szerezte meg. Lásd: Kashmir Hill: The Secretive Company That Might End Privacy as We Know It. *The New York Times,* 2020. január 18.

[22] Európai Bizottság: *Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyéb uniós jogalkotási aktusok módosításáról.* COM(2021)206 végleges (2021. április 21.).

[23] Európai Bizottság: *Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyéb uniós jogalkotási aktusok módosításáról.* COM(2021)206 végleges (2021. április 21.). (33) bevezető bekezdése.

különösen, hogy ezek az intézkedések biztosítsák, hogy a felhasználó a rendszerből származó azonosítás alapján ne hozzon intézkedést vagy döntést, kivéve, ha ezt legalább két természetes személy ellenőrizte és megerősítette.[24]

Deepfake – a mesterséges intelligencia használata az arcfelismerés megtévesztésére

„A nevem Werner Brandes. A hangom az útlevelem. Azonosítson engem." – hangzik az ismertté vált mondat a Komputerkémek (Sneakers) című 1992-es filmből, ahol a Robert Redford által alakított szakértő és csapata a célszemélytől származó hangfelvételekből összevágott, kicsit természetellenesen hangzó felvétellel végül sikeresen megtéveszti a csúcstechnológiás hangfelismerő rendszert.[25] Ma már nincs szükség a hagyományos kémtevékenység és az audiószerkesztés nagy rutint igénylő műveleteire ahhoz, hogy az eredetihez megtévesztésig hasonlító felvételt készítsünk bárkiről. A mesterséges intelligencia alapjain, különösen pedig a mély neurális hálók felhasználásával kialakult az úgynevezett *deepfake*-technológia. A kifejezés azt a módszert takarja, hogy a mesterséges intelligencia segítségével egy személyről rendelkezésre álló képek vagy hangfelvételek alapján olyan manipulált felvételt készítenek, amely a valós személyt egy olyan fiktív helyzetben ábrázolja, amilyen helyzetben a konkrét személy nem szerepelt, illetve olyan kijelentést tulajdonít neki, amilyen kijelentést ő nem tett meg.

A deepfake-felvételek már elérték azt a színvonalat, ahol egy embert könnyen megtévesztenek, mivel ha nincs mellette az eredeti hang (kép) referenciaként, akkor az agyunk hajlamos egy kicsit a valóságot is meghajlítani, és a hamis felvételt valóságosnak hinni. A deepfake-technológia azonban már nem csak az emberekre jelent veszélyt, a legmodernebb technológia, a GAN *(generative adversarial network)* felhasználásával olyan meggyőző hamisítványok készíthetők, amelyek a piacon elérhető arcfelismerő rendszereket akár 95%-os arányban is képesek megtéveszteni.[26] Az arcfelismerő rendszerek sérülékenységét általában nem a deepfake típusú támadásokkal kapcsolatban vizsgálják, a vonatkozó szakirodalomban először Tariq és munkatársai 2021-ben megjelent tanulmányában[27] olvashatunk arról, hogy empirikus vizsgálatnak vetették alá egyes nagy szolgáltatók arcfelismerő rendszereit a deepfake alapú támadásokat szimulálva. A szerzők arra a – nem túl megnyugtató – következtetésre

[24] Európai Bizottság: *Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyéb uniós jogalkotási aktusok módosításáról.* COM(2021)206 végleges (2021. április 21.). 14. cikk (5) bekezdése.

[25] Sneakers (1992): My Voice Is My Passport. *YouTube,* 2013. június 10.

[26] Pavel Korshunov – Sébastien Marcel: *DeepFakes: a New Threat to Face Recognition? Assessment and Detection.* (h. n.), (k. n.), 2018.

[27] Shahroz Tariq – Sowon Jeon – Simon S. Woo: *Am I a Real or Fake Celebrity? Evaluating Face Recognition and Verification APIs under Deepfake Impersonation Attac.* New York, Association for Computing Machinery, 2022.

jutottak, hogy az esetek döntő többségében a vizsgált arcfelismerő rendszereket meg lehetett téveszteni a deepfake-támadás útján.

A deepfake-felvételek elterjedése számos további negatív következményt hordoz magában, így az álhírek térnyerését és a hírekbe, illetve a látott vagy hallott felvételekbe való közbizalom jelentős erodálódását. Az álhír *(fake news)* – egyre inkább közkeletűvé váló – fogalma azt takarja, hogy a technológia segítségével manipulált módon, általában a közvélemény befolyásolása vagy megtévesztése céljából valótlan információkat vagy jelentősen elferdített tényeket közölnek valósként a nyilvánosság felé. A valós információkra megtévesztésig hasonlító álhírek térnyerésével párhuzamosan megfigyelhető jelenség az, hogy a nyilvánosság tagjai egyre inkább kétkedve fogadják a valóságos híreket is, így egy-egy nagyobb hatású hír kapcsán hamar elindul arról az eszmecsere, hogy az alapjául szolgáló felvételt esetleg manipulálták, így annak nem lehet hinni. Ezekre a társadalmi-politikai jelenségekre és következményeikre e tanulmány keretei között nem térünk ki, mivel azok általában nem kapcsolódnak biometrikus azonosító rendszerekhez, hanem a néző személy megtévesztésén alapulnak.

Az arckép elterjedése az online térben

A deepfake-technológia akkor működik hatásosan, ha kellő mennyiségű forrásanyag (kép, hang vagy videó) áll rendelkezésre ahhoz, hogy a rendszer megtanulhassa a célszemély jellemzőit utánozni. Ebben nagy segítségére vannak a közösségi platformok, kép- és videómegosztó oldalak, ahová a felhasználók számolatlanul töltik fel magukról a képeket. A kockázatot az hordozza, hogy ezeket a képeket és videókat az ismerősökön kívül idegenek, esetleg rosszindulatú személyek is könnyen megszerezhetik. A nagy közösségi oldalak viszonylag nehezen törhetők fel, azonban sok más lehetőség van a képek megszerzésére: a felhasználók nyilvános hozzáférésűre állított képeit és videóit az oldalak automatizált átfésülése és a talált állományok lemásolása *(scraping)* útján előre meg nem határozható külső szereplők tudják megszerezni.[28] Ráadásul harmadik felek által fejlesztett alkalmazások hozzáférhetnek különféle adatokhoz, kihozhatják azokat a zárt ökoszisztémából, hogy utána a saját rendszerükben, néha gyenge biztonság mellett tárolják. Emellett kikerülhet adat titkos adatmegosztó megállapodások vagy adatlopás útján is. Egy kutatás[29] informatikai csapdákkal mérte ezt az „elszivárgást", és arra jutott, hogy a vizsgált több mint ezer applikációból nagyjából húsz osztott meg adatot ismeretlen harmadik

[28] Matthew Doktor: Facial Recognition and the Fourth Amendment in the Wake of Carpenter v. United States. *University of Cincinnati Law Review,* 89. (2021), 2.

[29] Shehroze Farooqi – Maaz Musa – Zubair Shafiq – Fareed Zaffar: CanaryTrap: Detecting Data Misuse by Third-Party Apps on Online Social Networks. *Proceedings on Privacy Enhancing Technologies,* (2020), 4. 336–354.

féllel. Egyes alkalmazások (például a Facebook SDK és a Google AdSense) ehhez képest fordítva működnek: a külső applikáció oszt meg a nagy céggel olyan adatot, amelyet saját maga gyűjtött.

Az alacsony biztonsági szintű tárolás azonban nem csak a webes szolgáltatók által hordozott kockázat. Biometrikus adatokat (arckép, ujjlenyomat, hangminta, vénaszkenner által gyűjtött mintázat) számos kisebb-nagyobb szervezet gyűjt és tárol, így munkahelyek, iskolák, sportlétesítmények, de akár edzőtermek vagy közlekedési vállalatok is. Az adatok magas biztonsági szintet garantáló tárolása azonban jelentős költségráfordítással jár, és később is folyamatos odafigyelést igényel. Ezek a kisebb gazdasági erejű szereplők a tapasztalatok szerint ritkán fordítanak kellő gondot és anyagi áldozatot a biometrikus adatokat tároló rendszer adatbiztonsági intézkedéseire, így közvetetten kiteszik az ügyfeleik biometrikus adatait a rosszindulatú támadásoknak és adatgyűjtésnek. Egy hackercsoportnak jóval könnyebb dolga van ellopni a gyermekek biometrikus azonosító adatait egy kisváros általános iskolájának szerveréről, mint feltörni a Facebook vagy a Google dollármilliókból kiépített és fenntartott biztonsági rendszerét. A technológia praktikus alkalmazhatósága és relatív olcsósága sokszor arra csábítja ezeket a szervezeteket, hogy biometrikus azonosító rendszert építsenek ki, azonban azt kevesen gondolják végig, hogy a nem megfelelő adatbiztonsági intézkedések miatt ezáltal súlyos kockázatnak teszik ki saját ügyfeleik vagy munkatársaik szenzitív személyes adatait.

Új bűncselekményformák – az információs és pénzügyi rendszerek sebezhetősége

A fent bemutatott deepfake-technológia egy régóta létező bűncselekmény elkövetőinek kezébe ad egészen új eszközöket. Az az ötlet ugyanis nem új keletű, hogy valaki mást megszemélyesítve próbálunk hozzáférni az illető javaihoz vagy az általa kezelt értékekhez és információkhoz. Ennek kivédésére számos technológiai megoldás is a rendelkezésünkre áll már – többek között a biometrikus adatokon alapuló hitelesítés.

Néhány évvel ezelőtt adták hírül, hogy egy jelentős energetikai vállalat egyik vezetője nagy összegű banki utalást teljesített egy hackercsoport részére, akik a telefonbeszélgetés során a cégcsoport egyik igazgatójának hangját számítógépesen reprodukálva sikeresen tévesztették meg őt arról, hogy a főnöke utasítja a pénz elküldésére.[30] A bűnüldöző szervek meglepve álltak az eset előtt, mivel korábban még nem volt arra példa, hogy az emberi beszédhang hamisításával kövessenek el ilyen típusú bűncselekményt. A vállalati biztonsági mechanizmusokat is sikeresen kerülte ki az elkövető, mivel az azonosítást megkövetelő és a csalásellenes megoldások a digitális írott kommunikációra fókuszáltak, a telefonbeszélgetésekre nem.

[30] A 2019-es ügyről szóló beszámolóért lásd Catherine Stupp: Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal,* 2019. augusztus 30.

Az elkövetőket azóta sem sikerült elfogni, valamint az sem tisztázódott, hogy a telefonbeszélgetés során a támadók hogyan válaszoltak a vonal másik végén elhangzó kérdésekre, használtak-e mesterségesintelligencia-alapú csevegőalkalmazást *(chatbot)* erre. A beszédhang megtévesztésig hasonló klónozása már önmagában is nagy kockázatot hordoz, amit tovább fokoz az, ha valakinek a képi megjelenését is hamisítani tudják, így akár egy videókonferencia résztvevőit is meg tudják téveszteni.

A fentiekhez hozzá kell tenni azt is, hogy megvan annak a reális lehetősége, hogy valaki eltérítse a biometrikus azonosító rendszer bemeneti jelét, és a beépített érzékelő (például kamera) által felvett kép helyett a támadó által betöltött felvételt továbbítsa azonosításra az algoritmusnak. Így még a képfelvevő egységet sem kell fizikailag megközelítenie a támadónak.

A biometrikus adatok visszavonhatósága

Abban biztosak lehetünk, hogy a biometrikus azonosítók nem vonhatók vissza, vagy nem változtathatók meg úgy, mint egy aláírás, egy elveszett bélyegző vagy egy feltört jelszó. Ha valakinek az ujjlenyomata, hangmintája, vagy – jelen tanulmány témája szempontjából kiemelt helyen – az arcáról készült digitális lenyomat (képek, videók, szkennelt topográfia stb.) kikerül az ellenőrzése alól, akkor úgy tekinthetjük, hogy ez a hitelesítési mód véglegesen kompromittálódott. Nincs ugyanis arra észszerű esély, hogy a világhálóra kikerült vagy egy adatlopással érintett adatcsomagban megtalálható biometrikus azonosító minden egyes példányát véglegesen törölni tudjuk. Így bármikor felbukkanhat egy deepfake-módszerrel készített hamisítvány, amelyet valaki az érintett személy jogosultságainak feltörésére használhat.

A közösségi hálózatok működéséből fakadó jellegzetességek mellett figyelembe kell venni azt is, hogy ezeket a platformokat rendszeresen átfésülik, és az ott található adatokat lementik különböző szereplők. Ezt a legtöbb közösségi oldal kifejezetten tiltja a szabályzataiban, azonban technikai intézkedéssel kivédeni vagy kiszűrni nem tudja. Erre tekintettel nem túlzás azt állítani, hogy – a közösségi oldalaktól tudatosan távol maradó és a képmását semmilyen online felületen meg nem mutató kevesektől eltekintve – szinte minden ember arcmása, mozgása és hangja ki van téve a deepfake-támadásnak, és ezért nem tekinthető megbízható hitelesítő eszköznek. Következtetésként kimondhatjuk, hogy napjainkban, és különösen a közeli jövőben, a legkevésbé megbízható és hiteles azonosítási mód az arc- vagy hangfelismerés, ezért ezeket fokozott biztonságú rendszerekbe való belépéshez használni rendkívül kockázatos. Valószínűleg nem kevésbé kockázatos a saját technológiai eszközökbe való bejelentkezéshez használni ezeket, hiszen a legtöbben a mobileszközökön tárolják az üzleti és személyes életük adatkincseit, pénzügyi szolgáltatóik applikációit és ügyintézési felületeit. A kockázatok mérsékléséhez

vezető első lépés tehát annak tudatosítása lenne, hogy az arckép-, illetve hangalapú azonosítás egyáltalán nem megbízható módszer.

Jogi konzekvenciák, fejlesztési javaslatok

Az arc- vagy hangalapú azonosítás megbízhatatlansága, illetve a potenciális támadások elleni védekezés többféle következménnyel járhat. A hamisítás kockázatának mérséklése érdekében, a probléma gyökerére ható módon próbálkozhatunk megakadályozni vagy visszaszorítani az emberek arcképének, illetve a hang- és videófelvételeknek a nyilvános megosztását a közösségi hálózatokon. Ennek érdekében jelentős tudatosító és felvilágosító tevékenységre van szükség, ami felhívja a figyelmet a kép- és hanganyagok feltöltésének és megosztásának veszélyeire. Ennél szigorúbb fellépés lenne az, ha a jogalkotó kifejezetten megtiltaná vagy korlátozná az ilyen szolgáltatások működését. A jelenleg látható technológiai és társadalmi tendenciák azonban éppen ezzel ellentétes irányban hatnak, a kép- és videóalapú platformok népszerűsége egyre növekszik, és az általuk nyújtott szolgáltatások a mindennapok részévé váltak. A szabályozó hatóságnak minden bizonnyal el kell fogadnia azt a tényt, hogy a személyek képmása megállíthatatlanul terjed a világhálón, és hamarosan szinte mindenkiről rendelkezésre fog állni elegendő forrásanyag a jó minőségű deepfake-videók vagy hanganyagok elkészítéséhez. Ennek elismerése elvezet minket oda, hogy a megelőzés helyett a probléma kezelésére és az azzal való hosszú távú együttélésre kell kialakítani a szabályrendszert. Az elképzelhető megoldások egy része a deepfake-tartalmak készítésének tiltására vagy azok megjelölésére vonatkozik, másik részük a biometrikus arcfelismerő rendszerek működésének szabályozására. A javaslatok harmadik iránya pedig az eredeti felvételek hitelességének biztosítása által próbálkozik kiszűrni a hamisítványokat.

A deepfake tiltása

A tiltás vagy büntetőjogi szankció kilátásba helyezése a jogban általában *ultima ratio* jelleggel alkalmazható eszköz, mi azonban mégis elsőként tárgyaljuk. Ennek oka az, hogy valószínűleg ez a legkevésbé alkalmazható megoldás mind közül. A deepfake-technológia használatának, illetve az erre használható alkalmazások fejlesztésének vagy elérhetővé tételének tiltása ugyanis rendkívül túlzó beavatkozás, és nincs tekintettel arra, hogy ennek a technológiának zömében társadalmilag hasznos, illetve kárt vagy kockázatot nem okozó felhasználási területei vannak. A technológia legnagyobb potenciális felhasználója a szórakoztatóipar: hollywoodi filmstúdiók a drága CGI vagy a kicsit elavult *greenbox* technológia helyett tudják használni a színészeikről készített deepfake-felvételt, amelynek a segítségével ki tudják javítani a felvételkor észre nem vehető hibákat, szükség esetén utólag módosíthatnak a dialóguson, és nem kell nagy költségen egy megismételt forgatási napot

tartani ezek miatt az apró korrekciók miatt.[31] Ugyanígy alkalmas a technológia a színészek arcának fiatalítására vagy öregítésére, ahogy a 2019-es év sikerfilmje, az *Irishman* esetében már láttuk. A deepfake-technológiát használja a reklámipar is annak érdekében, hogy egy híresség által elmondottakat minél több nyelven tudják élethűen célba juttatni.[32] Az oknyomozó újságírók és a dokumentumfilmesek is örülnek ennek a lehetőségnek, mert ezáltal képesek megváltoztatni a védett vagy veszélynek kitett forrásaik arcát és hangját úgy, hogy továbbra is szerepeltetik őket a képernyőn. A deepfake-módszerekkel előállított hang- és képanyagok megjelennek az oktatás területén is: a Covid-járvány által is felerősítve egyre inkább teret nyernek a digitális oktatási formák. Az oktatók számára kiváló lehetőséget jelent az, hogy az írott tananyag hangoskönyvvé alakításához nem kell neki felmondani az egész tankönyvet vagy a megírt előadásszöveget, hanem egy szövegolvasó szoftver segítségével – akár a tanár saját hangján megszólaló – hanganyaggá lehet alakítani az írott szöveget. Ez a látássérült emberek mindennapi életét is megkönnyíti, az írott szövegek hanggá formálása által. A deepfake-technológia ebben az esetben hozzájárul ahhoz, hogy az emberi beszédhez jobban hasonlító intonációjú és beszédtempójú hanganyag készülhessen el, elkerülve a „robotszerű" gépi hang monotonitását.

Összességében tehát a deepfake tiltása számos pozitív hatású felhasználási lehetőségtől is elzárná a társadalmat. Emellett pedig egy ilyen tiltás igen nehezen kivitelezhető, hiszen a technológia viszonylag egyszerűen, olcsón, bárki számára hozzáférhető módon megjelent már a világhálón, így annak a terjedését visszaszorítani nem életszerű.

A deepfake megjelölése

Megoldás lehet a tiltás helyett az, ha a jogalkotó előírja a tartalomkészítők számára a deepfake-tartalmak világos és jól látható megjelölését. Ezáltal a médiatartalom nézője vagy hallgatója nem esik tévedésbe a felvétel valódisága felől, hanem egyértelmű lesz számára az, hogy amit lát, nem valós felvétel, annak szereplője valójában nem tette vagy mondta azt, amit a felvétel bemutat. Ezen a nyomon indult el az Egyesült Államok kongresszusa, amely egy 2018-ban bizottsági szinten elbukott javaslat után 2019-ben átfogó törvényjavaslatot vett napirendjére ebben a témában: ez volt a DEEP FAKES törvény (Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act

[31] Cooper Hood: How Deepfake Technology Can Change The Movie Industry. *Screenrant,* 2021. augusztus 29.

[32] Így például a David Beckham részvételével készült társadalmi célú kampányban a futballsztár kilenc nyelven mondja el a szövegét. Lásd: David Beckham Speaks Nine Languages to Launch Malaria Must Die Voice Petition. *YouTube,* 2019. április 9.

of 2019).[33] A törvényjavaslat a deepfake-médiatartalmak készítői számára legalább egyszeri szóbeli és a képfelvétel teljes időtartama alatt látható vizuális figyelemfelhívás megjelenítését írja elő. A megjelölési kötelezettség megsértése vagy a jelölés utólagos eltávolítása miatt pedig öt évig terjedő szabadságvesztés büntetést helyez kilátásba, és minden egyes ilyen felvétel után 150 ezer dollárig terjedő magánjogi felelősséget állapít meg. A törvényjavaslat előremutató eleme emellett, hogy a csalás és hamis megszemélyesítés büntetőjogi tényállását kiterjeszti a deepfake-felvételek útján való elkövetésre is. A javaslat végül nem emelkedett törvényerőre, azonban fontos iránymutatásul szolgál a jövőbeni jogalkotási próbálkozások számára.

Az Európai Unió mesterséges intelligenciáról szóló rendelettervezete hasonlóképpen előírja, hogy „azon MI-rendszerek felhasználói, amelyek olyan, meglévő személyekre, tárgyakra, helyekre vagy más szervezetekre vagy eseményekre érzékelhetően hasonlító képet, audio- vagy videotartalmat generálnak vagy manipulálnak, amely egy személy számára megtévesztő módon eredetinek vagy valóságosnak tűnhet *(„deepfake"),* közlik, hogy a tartalmat mesterségesen hozták létre vagy manipulálták".[34]

A megjelölés tehát járható út lehet arra a célra, hogy a deepfake-videók, -képek vagy -hangok ne kerülhessenek véletlenül egy biometrikus azonosító rendszerbe, ahol azokat egy valós személy hitelesítésére vagy felismerésére használják. Ilyen esetekben az azonosító rendszer vagy az azt felügyelő felhasználó detektálja a vízjelet, és nem használja ezt a felvételt a továbbiakban. Mindazonáltal a megjelölési kötelezettség csak az őszinte szándékú felhasználókat tartja távol az ilyen magatartástól, a rosszindulatú támadókra a jogszabályi kötelezés vagy tiltás nincs hatással.

A biometrikus azonosító rendszerek működésének korlátozása

A biometrikus azonosító rendszerek fejlesztői és a deepfake-hamisítványok előállítói között kiélezett fegyverkezési verseny zajlik. A korábban már bemutatott amerikai törvényjavaslat a nemzetbiztonsági minisztérium (DHS) égisze alatt kívánta létrehozni a deepfake-médiatartalmak felismerését kutató részleget, emellett több gazdasági szereplő is foglalkozik ilyen megoldások fejlesztésével.[35] Számos technológiai megoldás létezik már a hamisított bejelentkezési adatok kiszűrésére, azonban ezzel együtt folyamatosan javul a deepfake-tartalmak minősége

[33] H.R.3230 — 116th Congress (2019–2020).

[34] Európai Bizottság: *Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról.* COM(2021)206 végleges (2021. április 21.). 52. cikk (3) bekezdése.

[35] When Seeing is no Longer Believing. Inside the Pentagon's Race against Deepfake Videos. *CNN,* (é. n.).

is. Meglátásom szerint a megoldást az jelentené, ha a távoli vagy jelenlétet igénylő biometrikus azonosító rendszerek használatát felváltaná más, nehezebben megszerezhető vagy hamisítható azonosítási mód. A két- vagy többfaktoros hitelesítés szintén megoldást jelenthet erre a problémára, ezért érdemes szorgalmazni annak kötelezővé tételét minden olyan esetben, ahol kép- vagy hangalapú hitelesítés történik. Emellett a gazdasági szférában vagy a magánéletben használt biometrikus azonosító rendszerek esetében a rendszer integritását, a külső behatolás vagy tartalombehelyezés észlelhetőségét folyamatosan biztosítani és rendszeresen ellenőrizni kell.

A rendvédelmi szervek által használt távoli azonosító rendszerek esetében a hibás azonosítás – a fent már bemutatott okok miatt – még súlyosabb következményekhez vezethet.[36] Ezért az ilyen rendszerek esetében – az EU rendelettervezetében foglaltakhoz hasonló módon – kötelező érvénnyel rögzíteni kell, hogy önmagában a biometrikus azonosító rendszer eredménye alapján a hatóságok nem tehetnek az egyén jogait vagy szabadságát is érintő lépéseket, azt valamilyen más azonosító tényezővel (például mobilhelyadatok, helyszínen rögzített nyomok) előtte meg kell erősíteni. A biometrikus távoli azonosító rendszerek által adott eredmények bíróság előtti felhasználása során figyelembe kell venni a rendszer által adott eredmény bizonyossági (pontossági) fokát is, valamint biztosítani kell az így keletkezett bizonyíték vitathatóságát.[37] Mindazonáltal érdemes itt felhívni a figyelmet arra is, hogy az ember mint szemtanú legtöbbször pontatlanabb az azonosításban és felismerésben,[38] illetve maga is ki van téve számos elfogultsági tényezőnek.[39]

Az eredeti felvételek integritásának biztosítása

A hamisítványok megjelölése a nyilvánosság tájékoztatása esetében jó eszköz lehet a félreértések elkerülésére. Ez azonban nem szűri ki a rosszindulatú aktorokat, hiszen ők eleve jogsértő cselekményt kívántak végrehajtani. A technológia mai állása azonban lehetőséget ad ennek a gondolatnak a megfordítására is: elő lehet írni az „eredeti" felvételek megjelölésének kötelezettségét és azok integritásának ellenőrizhetőségét. A digitális hang-, kép- vagy videófelvétel készítését követően

[36] Rachel S. Fleischer: Bias in, Bias out: Why Legislation Placing Requirements on the Procurement of Commercialized Facial Recognition Technology Must Be Passed to Protect People of Color. *Public Contract Law Journal,* 50. (2020), 1. 65.

[37] Ari B. Rubin: A Facial Challenge: Facial Recognition Technology and the Carpenter Doctrine. *Richmond Journal of Law & Technology,* 27. (2021), 2. 24–34.

[38] Hal Arkowitz – Scott O. Lilienfeld: Why Science Tells Us Not to Rely on Eyewitness Accounts. *Scientific American,* 2010. január 1.

[39] Mollie McGuire – Kathy Pezdek: Birds of a Feather get Misidentified Together: High Entitativity Decreases Recognition Accuracy for Groups of Other-Race Faces. *Legal and Criminological Psychology,* 21. (2016), 1. 202–211.

a felvevőkészülék a mentett fájlról egy biztonságos és titkosított matematikai lenyomatot (úgynevezett *hash*) tud készíteni. Ez a kulcs az adott médiatartalomra jellemző, egyedi azonosító lesz. Ha később ellenőrizni szeretnénk azt, hogy nálunk az eredeti fájl vagy egy módosított változat van-e, nincs más tennivalónk, mint elkészíteni a saját fájlunk hash lenyomatát és összevetni az eredetivel. Ha a dokumentumon bármilyen módosítást eszközöltek, akkor a két hash-érték nem fog megegyezni. Így gyorsan ellenőrizhető az, hogy a felvevőkészülékből származó eredeti tartalmat látjuk, vagy pedig egy módosított változatot.

A hash-érték generálása és tárolása mellett egy másik modern technológiát, a blokkláncokat is segítségül hívhatjuk. Egy felvétel első nyilvános közzététele vagy elkészítése esetén a felvétel hash-értékét blokkláncon tárolhatjuk, amely a felvétel minden későbbi módosítását (akár egyszerű képszerkesztési vagy vágási műveleteket) megmásíthatatlan és kitörölhetetlen módon rögzíti és tárolja. Így elvileg nyomon követővé válik az, hogy ki (vagy mely eszköz), mikor és milyen módosításokat hajtott végre a felvételen. Ha a blokklánc nem folytonos, sérült vagy nem érhető el, akkor gyanakodhatunk a manipulációra.

A fenti két megoldás együtt hozzájárulhat a felvételek hitelességébe vetett közbizalom megőrzéséhez, elősegíti a hamisítványok kiszűrését és a visszaélések megakadályozását. Ezek alkalmazásához arra van szükség, hogy a jogalkotó a hang- és képfelvételek készítésére alkalmas eszközök gyártóival vagy forgalmazóival szemben előírja ezeknek a műszaki megoldásoknak a beépítésére való kötelezettséget. Így néhány év alatt a lakosság és az üzleti felhasználók birtokában már olyan eszközök lesznek, amelyek maguk hitelesítik az általuk készített felvételeket.

Összegzés

A biometrikus – és különösen a hang-, illetve képalapú – azonosító megoldások egyre szélesebb körben nyernek teret a mindennapjainkban, az üzleti felhasználástól a szórakoztató elektronikán át a rendvédelmi szervek általi használatig. Informatikai szempontból a felhasználó azonosítására kifejezetten megbízható megoldásként szokták feltételezni ezeket, azonban a közelmúlt társadalmi és műszaki fejlődése ezt megkérdőjelezi. A kép- és hangfelvételek hamisítására használható, mesterséges-intelligencia-alapú úgynevezett deepfake-technológia lehetővé teszi azt, hogy kellő mennyiségű forrásanyag (az érintett személyről készült felvétel) rendelkezésre állása esetén nagyon meggyőző hamisítványokat készíthessünk. Tekintettel arra, hogy a közösségi portálok, kép- és videómegosztó oldalak felhasználói saját maguk szolgáltatják nagy mennyiségben ezt az elérhető nyersanyagot, azt kell feltételeznünk, hogy mára már szinte bárkiről készíthető deepfake-felvétel. Ezek segítségével pedig megtéveszthetők a biometrikus azonosító rendszerek, aminek – attól függően, hogy mire használják az azonosító rendszert – akár súlyos egyéni és társadalmi következményei is lehetnek.

A tanulmányban körüljártuk a megoldásra nyitva álló lehetőségeket és a jelenleg létező vagy tervezett jogalkotói válaszlépéseket. Arra jutottunk, hogy a deepfake-tartalmak megjelölése csak annyiban hasznos, amíg annak előállítója jó szándékú, jogkövető magatartást tanúsít. A rosszindulatú aktorok ellen egyfelől az azonosító rendszerek megerősítésével, illetve korlátozott használhatóságuk és kockázatos mivoltuk elismerésével védekezhetünk. Másfelől pedig bemutattuk azt a koncepciót, amely a jelenlegi – a hamisítványok megjelölésére koncentráló – jogalkotói gondolkodásmódot megfordítja, és az eredeti felvételek megjelölését és nyomonkövethetőségét írja elő. Ez a megoldás, a felvételt készítő hardvereszközökbe kódolva, hozzájárulhat a rosszindulatú manipuláció elleni védekezéshez.

"It's Time to Change It" – The Fundamental Rights Risks of Using Facial Recognition Systems in the Age of Social Media and Deepfake

Abstract

Recent advances in technology have made it possible to create more accurate remote facial recognition systems than ever before. We encounter these systems every day: this is how our smartphone recognises us, we enter our workplace using this method, and there is such an algorithm behind the surveillance camera systems on the streets. In this study, we draw attention to the fact that, although it is getting more and more popular, the use of a facial recognition system involves a number of legal risks; using the large amount of imagery available from someone on the World Wide Web, it became possible to create fake images of a person, which can deceive even advanced facial recognition systems. Therefore, in this study, we propose solutions to regulate the operation of facial recognition technology and to effectively combat the abuse of facial images.

Felhasznált irodalom

Apple: *A Face ID használata iPhone és iPad Pro készüléken.* Online: https://support.apple.com/hu-hu/HT208109

Arkowitz, Hal – Scott O. Lilienfeld: Why Science Tells Us Not to Rely on Eyewitness Accounts. *Scientific American,* 2010. január 1. Online: https://bit.ly/3oasecQ

David Beckham Speaks Nine Languages to Launch Malaria Must Die Voice Petition. *YouTube,* 2019. április 9. Online: www.youtube.com/watch?v=QiiSAvKJIHo

Doktor, Matthew: Facial Recognition and the Fourth Amendment in the Wake of Carpenter v. United States. *University of Cincinnati Law Review,* 89. (2021), 2.

Európai Bizottság: *Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról.* COM(2021)206 végleges (2021. április 21.).

Facial Recognition – Frequently Asked Questions. *State of Michigan,* 2022. június. Online: https://bit.ly/3AXsrrx

Farooqi, Shehroze – Maaz Musa – Zubair Shafiq – Fareed Zaffar: CanaryTrap: Detecting Data Misuse by Third-Party Apps on Online Social Networks. *Proceedings on Privacy Enhancing Technologies,* (2020), 4. 336–354. Online: https://doi.org/10.2478/popets-2020-0076

Fleischer, Rachel S.: Bias in, Bias out: Why Legislation Placing Requirements on the Procurement of Commercialized Facial Recognition Technology Must Be Passed to Protect People of Color. *Public Contract Law Journal,* 50. (2020), 1. 63–89.

Frontex: *Best Practice Technical Guidelines for Automated Border Control (ABC) Systems.* 2015. augusztus 31.

Grother, Patrick – Mei Ngan – Kayee Hanaoka: *Face Recognition Vendor Test (Frvt) Part 3: Demographic Effects.* National Institute of Standards and Technology, U.S. Department of Commerce, Nistir 8280, 2019. december. Online: https://doi.org/10.6028/NIST.IR.8280

Hill, Kashmir: Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match. *The New York Times,* 2020. december 29. Online: https://nyti.ms/3AYCeO2

Hill, Kashmir: The Secretive Company That Might End Privacy as We Know It. *The New York Times,* 2020. január 18. Online: https://nyti.ms/3ca7Y8q

Hill, Kashmir: Wrongfully Accused by an Algorithm. *The New York Times,* 2020. június 24. Online: https://nyti.ms/3REDN9Q

Hood, Cooper: How Deepfake Technology Can Change The Movie Industry. *Screenrant,* 2021. augusztus 29. Online: https://bit.ly/3RIYxwY

Hughes, Neil: A Backlash Against Facial Recognition is Brewing. *TechHQ,* 2019. augusztus 20. Online: https://bit.ly/3cpNWqV

Korshunov, Pavel – Sébastien Marcel: *DeepFakes: a New Threat to Face Recognition? Assessment and Detection.* (h. n.), (k. n.), 2018. Online: https://arxiv.org/pdf/1812.08685.pdf

McGuire, Mollie – Kathy Pezdek: Birds of a Feather get Misidentified Together: High Entitativity Decreases Recognition Accuracy for Groups of Other-Race Faces. *Legal and Criminological Psychology,* 21. (2016), 1. 202–211. Online: https://doi.org/10.1111/lcrp.12066

Meyer, Bernard: Most Common Passwords: Latest 2022 Statistics. *Cybernews,* 2022. május 27. Online: https://bit.ly/3aJVRid

Mi az az arcfelismerési beállítás a Facebookon, és hogyan működik? *Facebook.* Online: https://bit.ly/3yQIUv7

Rubin, Ari B.: A Facial Challenge: Facial Recognition Technology and the Carpenter Doctrine. *Richmond Journal of Law & Technology,* 27. (2021), 2. 24–34.

Samsung: *What are the Requirements for Using Face Recognition in Samsung H Series TV(UA40H6400)?* 2020. október 14. Online: https://bit.ly/3yLA6X5

Sanchez del Rio, Jose – Daniela Moctezuma – Cristina Conde – Isaac Martin de Diego – Enrique Cabello: Automated Border Control E-Gates and Facial Recognition Systems. *Computers & Security,* 62. (2016). 49–72. Online: https://doi.org/10.1016/j.cose.2016.07.001

Sauer, Pjotr: Privacy Fears as Moscow Metro Rolls out Facial Recognition Pay System. *The Guardian,* 2021. október 15. Online: https://bit.ly/3Pe0P5V

Sheehan, Michael J. – Michael W. Nachman: Morphological and Population Genomic Evidence that Human Faces Have Evolved to Signal Individual Identity. *Nature Communications,* (2014), 5. Online: https://doi.org/10.1038/ncomms5800

Sneakers (1992): *My Voice Is My Passport.* YouTube, 2013. június 10. Online: www.youtube.com/watch?v=-zVgWpVXb64

Street, Francesca: How Facial Recognition is Taking over Airports. *CNN Travel*, 2019. október 8. Online: https://cnn.it/3zaDX1w

Stupp, Catherine: Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*, 2019. augusztus 30. Online: https://on.wsj.com/3INVCPS

Tariq, Shahroz – Sowon Jeon – Simon S. Woo: *Am I a Real or Fake Celebrity? Am I a Real or Fake Celebrity? Evaluating Face Recognition and Verification APIs under Deepfake Impersonation Attack*. New York, Association for Computing Machinery, 2022. Online: https://doi.org/10.1145/3485447.3512212

Watts, Marina: Read This Chart to Find Out How Fast Someone Can Hack Into All of Your Accounts. *Newsweek*, 2020. január 9. Online: https://bit.ly/3RGRcxX

When Seeing is no Longer Believing. Inside the Pentagon's Race Against Deepfake Videos. *CNN*, (é. n.). Online: https://cnn.it/3cm26Jw