JEL-KÉP

KOMMUNIKÁCIÓ, KÖZVÉLEMÉNY, MÉDIA

2022/3. szám

DEEPFAKE ÚTIKALAUZ LAIKUSOKNAK, AVAGY BEVEZETÉS AZ AUDIOVIZUÁLIS MÉDIAMANIPULÁCIÓBA

Merziczky Marcell

Christian Rathgeb – Ruben Tolosana – Ruben Vera-Rodriguez – Christoph Busch (szerk.) *Handbook of Digital Face Manipulation and Detection. From DeepFakes to Morphing Attacks*. Springer, 2022.

A multitasking korát éljük. Nagyon kevés olyan dolog van, amit nem tudunk egy másik cselekvéssel párhuzamosan végezni, ez pedig megalapozza figyelmünk széttöredezését. Egyre nehezebben koncentrálunk kizárólag egy ingerre, és az a koncentráció sem tart sokáig. Ez a folyamat a kommunikációt sem kerüli el: egyszerűbb üzeneteket kell megkapóbb módon küldenünk, ha azt akarjuk, hogy a másik fél rászánja a kognitív erőfeszítést a befogadásra. A hangsúly fokozatosan tolódik el a vizualitás felé a verbalitás ellenében, agyunk ugyanis a képeket gyorsabban dolgozza fel a(z írott) szöveghez képest. A problémát az jelenti, hogy hiába vagyunk tisztában azzal, hogy a közösségi média platformok elterjedésével korlátlan lehetőségünk lett a képek-videók manipulálására is, az információk igazságértékét ezzel együtt is arcokhoz kapcsoljuk. A mélytanuláson alapuló kép- és videómanipuláció – a deepfake – pedig pontosan ezt használja ki.

A Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez és Christoph Busch szerkesztette kötet, a *Handbook of Digital Face Manipulation and Detection – From DeepFakes to Morphing Attacks* az imént taglalt okokból kifolyólag arra vállalkozik, hogy a laikus médiafogyasztóknak útmutatást adjon a vizuális manipulációs módszerekben, amelyekkel mindannyian szembesülhetünk mindennapi internethasználatunk közben, és amelyek bővebb ismerete máris könnyebbé teheti a megtévesztés ellen való védekezést. Az olvasó betekintést kaphat azokba a technológiai folyamatokba, amelyek a deepfake videók előállítása mögött húzódnak meg, felzárkózhat a tudományterülethez tartozó kulcsszavakból (mint amilyen a GAN, az arcfúzió vagy a szintetikus személy) és megismerheti a legkorszerűbb módszereket a hamisítás detektálására.

Olyan tudományterületek konvergálódnak ebben a kézikönyvben, mint a biometria, a kiberbiztonság, a multimédia kriminalisztika, a gépi látás és a szociológia. A mű nem titkolt célja, hogy hiánypótló módon összegyűjtse az egyes témák szakértőinek munkáját, a lehető legtöbb irányból megközelítve a mélyhamisítás jelenségét annak érdekében, hogy átfogó gondolati struktúrákkal gazdagítsa a befogadók világképét.

A Springer kiadó gondozásában közreadott olvasmány tulajdonképpen egy tanulmánykötet, amely négy nagy szakaszra bontható: az első fókuszában – egyfajta bevezető jelleggel – magának a kép- és videómanipulációs eljárásoknak a bemutatása áll, és ezek tágabb értelemben vett hatásai (például az arcfelismerő-rendszerek működésének hatékonyságára, sebezhetőségére); a második középpontjában manipulált arcokat tartalmazó kontent gyakorlati létrehozása, generálása; a harmadikban pedig a jelenlegi kiszűrési lehetőségeik. Végül, a negyedik részben jut elsődleges szerephez a társadalomra gyakorolt hatás és az etikai aspektus. A kötet a megismert módszerek összegzésével, a limitációkkal és a jövőbeli kutatások kiindulási pontjául szolgáló kérdésfelvetésekkel zárul.

Semmit a szemnek – a legnépszerűbb arcmanipulációs eljárások

A deepfake 2017 óta van jelen életünkben. Alapjául a mélytanulás folyamata szolgál, amely lehetővé teszi a felhasználók számára, hogy egy képen vagy videón szereplő személy arcát egy másik személy arcára cseréljék, digitális úton manipulálják. A szerzők kiemelik: a terület fokozatos gyarapodása elvezetett odáig, hogy napjainkra már szinte gyerekjáték nemlétező arcokat létrehozni, vagy a már meglévőket manipulálni, akár álló-, akár mozgóképeket tekintve. Ennek fő oka a mélytanuló algoritmusok fejlődése mellett az, hogy a szükséges (vizuális) adatok könnyen hozzáférhetők, bárki által összegyűjthetők az interneten. Létrejöttek kifejezetten erre specializálódó, nyilvános telefonos applikációk (mint amilyen a FaceApp), ezeket pedig előzetes tudás nélkül is egyszerű használni, akár már egy szelfivel is élethű végeredmény generálható. Az arccsere számos almódszert foglal magába, a kötet első szekciójában ezek közül mutatnak be a szerzők hatot, amelyek az utóbbi években a legnagyobb hatást gyakorolták a befogadókra: 1. teljesen mesterséges arc létrehozása (*entire face synthesis*), 2. személyazonosságcsere (*identity swap*), 3. arcfúzió (*face morphing*), 4. testi jellegzetesség manipulációja (*attribute manipulation*), 5. arckifejezés cseréje (például arcrekonstrukció, *expression swap*), 6. hang és/vagy szöveg alapján történő videóalkotás (*audio- and text-to-video*).

A növekvő számú, módosított arcokat ábrázoló audiovizuális tartalom a biometrikus rendszerek működésére is nagy terhet ró, hiszen ezek elsőszámú célja, hogy biológiai és viselkedésbeli minták (pl. ujjlenyomat, arc vagy az írisz) alapján ismerje fel az individuumokat. A hasonló elven működő programok döntéshozatali folyamataiba való betekintés mellett a szerzők ezek hatékonyságát is tesztelik tanulmányukban, aminek eredményeként kijelenthető, hogy sajnos nem tekinthetők biztonságosnak az arccserélős és a több arcot egyesítő technikákkal szemben, és a tanulmányok eredményeinek felhasználásával fejlesztéseket sürgetnek. Az első részben emellett a multimédia kriminalisztika deepfake előtti működési módjaiba is betekintést engednek.

A gép forog – a digitális arcmanipuláció folyamata és detektálása

A könyv második és harmadik szekciója a gyakorlatban mutatja meg azokat az eljárásokat, amelyeket a korábbi fejezetekben megismerhettünk. A fejezetek emellett számos olyan adatbázist tárnak az olvasó elé, amelyek kifejezetten deepfake technológiával megalkotott, mesterséges arcokat tartalmaznak. Ezek közül is kiemelik azokat, amelyek a Google-höz és a Facebookhoz kapcsolódnak, valamint több esetben maguk a tanulmányok szerzői biztosítanak újabb gyűjteményeket a műveletek elvégzésére. Rámutatnak ezek hibáira, így hívva fel a figyelmet egy kevésbé taglalt, ám annál lényegesebb pontra: a szűrőprogramok egy része szintén a mélytanuláson alapuló adatelemzést használja, viszont ha a tanuló adatbázis hibás vagy kevésbé jóminőségű felvételekből áll, akkor az algoritmus is rosszul szabályszerűsíti a számára betáplált műveleteket. Másként fogalmazva, a gyenge minőségű felvételekkel edzett program úgy viselkedik majd, mint egy biztonsági őr, akinek nem elég éles a látása: nem fogja tudni megkü-

lönböztetni a hamisítványt a valóditól. A felismerő algoritmusok közül három főbb kategóriát emelnek ki: a fizikai/pszichológiai tulajdonságokon alapulókat – ezek leplezik le többek között a pislogás furcsaságait vagy az összefüggéstelen fejmozgást; az arc jellegzetességein alapulókat; és az adaton alapulókat, amelyek kifejezetten valós és hamisított videókon sajátítják el az árulkodó jeleket.

A második és harmadik rész érezhetően perspektívát vált az olvasóbarát, laikusoknak szóló első részhez képest, ugyanis számos, nehezen lefordítható idegen kifejezést, mozaikszót tartalmaz. Itt érezhető a legjobban annak a hátránya, hogy nem egy saját ívvel rendelkező művel van dolga az olvasónak, hanem csupán azonos téma szerint összeválogatott tanulmányok gyűjteményével: túlságosan egy sémára épülnek a fejezetek alapjául szolgáló cikkek: bemutatnak egy módszert, amelyet tesztelnek különböző adathalmazokon, és rövid következtetést vonnak le az eredményességet illetően. A megértést továbbá nehezítik azok a képletek, amelyekkel a folyamatokat leírják, mintegy matematikai módon, illetve az eredmények szemléltetésére használt grafikonok és diagramok is. Utóbbiak különösen komplikáltan értelmezhetők anélkül a technológiai háttértudás nélkül, amelyek a mélytanulás és a gépi adatelemzés területeihez kapcsolódnak. Ezzel valószínűleg a könyv szerkesztői is tisztában vannak, az alkotás előszavában ugyanis úgy jellemzik az említett részeket, mint amik mélyebb elmerülést kínálnak a haladó olvasók számára.

Szociológiai és etikai kihívások, kitekintés

A kötet negyedik része bizonyos értelemben folytatja a második és harmadik részben megkezdett gondolatmenetet, hiszen foglalkozik még a különböző biometrikus rendszereket megtévesztő és a külső megjelenést módosító eljárások detektálásával, azonban szélesebb perspektívába helyezi az audiovizuális manipulációt. A hangsúly immár azokra a társadalmi változásokra kerül, amelyek elhozzák a „poszt-digitális világot” („*post-digital world*”), azaz a kort, melyben a digitális megoldások szorosan összefonódnak mindennapi életünkkel, attól elválaszthatatlanok lesznek. A szerzők „negyedik forradalomként” írják le többek között a mesterséges intelligencia felemelkedését, a virtuális valóság létrejöttét, a dolgok internetét („*Internet of Things*” – IoT) vagy a big data elemzési folyamatait – ezek a szegmensei világunknak néhány évtizeddel ezelőtt csupán sci-fi filmek részleteinek tűntek, ma pedig kézzelfogható valóságként vannak jelen. A technológia fejlődésével pedig a kihívások is új szintre emelkednek: a személyazonosságunk távoli azonosításának lehetősége a koronavírus-járvány következtében kialakult online ügyintézés folyamatában kulcsfontosságúvá vált, ezzel együtt azonban veszélyessé is, hiszen a biometrikus adatok meghamisítása ugyanúgy lehetőséggé lépett elő. Egy gép ellopná személyazonosságunkat? Néhány éve legyintettünk volna a felvetésre, ma azonban valósággá vált.

Az utolsó nagy fejezetül szolgáló cikkben ezzel összefüggésben a szerzők kifejtik, milyen limitációi és nehézségei vannak az egyes felismerési rendszerek pontosságára irányuló törekvéseknek, illetve, mik azok a megpróbáltatások, amelyeket maguk után vonnak. Az algoritmikus megoldások mellett az emberi felismerés is természetesen egy módja lehet a manipulált tartalmak elleni védekezésnek, azonban annak jelenlegi hatékonysága még a gépi szintet sem éri el: a fejezetben idézett kutatásból kiderül, hogy a nem jártas megfigyelők csupán 50%-ban ismerik fel a manipulált tartalmakat (tehát szinte egy érme feldobásával egyenlő statisztikai eséllyel), ez az arány az edzett megfigyelőknél is csak 60%. Ráadásul a szintetikus, tehát teljesen számítógép alkotta arcokat élethűbbnek gondolták a résztvevők (68%-os arányban, ezzel szemben a valódi arcokat csak 52%-ban). Nem elhanyagolható azonban az az aspektus, hogy míg számos kihívás és nehézség vár az emberiségre a deepfake és a hozzá hasonló technológia tükrében, több területen kifejezetten áldás lehet az arccserélős technológiák alkalmazása. Ilyen példának okáért a filmkészítés (pl. animációs munkálatok költséghatéko-

nyabbá tétele), az e-kereskedelem (pl. virtuális próbafülke funkcióval), az e-learning (pl. történelmi alakok életre keltésével hozni közelebb a különböző témákat a diákokhoz) és az e-egészségügy is (pl. virtuális személyek használata terápiás célokra).

A több nagy részből álló, rengeteg szerző munkáját összesítő kézikönyv útmutatást ad olvasójának a területhez, amelyről túlzás nélkül kijelenthető, hogy nagy mértékben lesz felelős jövőnk formálásáért. A digitális médiamanipulációs módszerekben és detektálási lehetőségeikben való elmerülés azonban nem egy könnyed csobbanás: az első és az utolsó fejezeteket leszámítva sokszor bonyolultan és egyirányúan, technológiai perspektívából mutatja be a témaköröket, amelyek így nem teljesen a laikus, hanem a bizonyos mértékben edzett befogadóknak kínálnak szélesebbkörű ismeretanyagot. A hiánypótló jelző azonban helytálló: a jelen kötet létrejöttéhez szükséges összefogás és közös munka lesz szükséges ahhoz is, hogy időtálló válaszokat tudjunk adni azokra a kérdésekre, amelyek a fejezetek olvasása közben merülnek fel.