

NEM HISZÜNK A

SZEMÜNKNEK?

GENERATÍV MI ÉS A DIGITÁLIS BIZONYÍTÁS VÁLSÁGA



A generatív mesterséges intelligencia nem egyszerűen új tartalmakat hoz létre, hanem alapjaiban írja át azt, ahogyan a valóságot felismerjük és értelmezzük. A képek és hangok egyre könnyebben és gyakrabban elszakadnak az eredetüktől, a hitelesség többé nem a tartalomban rejlik, hanem azokban a rendszerekben, amelyek képesek az eredetüket igazolni. A kérdés így lassan nem az, hogy mi igaz, hanem az, hogy ki és milyen feltételek között teszi azt hihetővé.





**Már nem az a kérdés,
hogyan manipulálható-e a valóság, hanem az,
hogyan egyáltalán felismerhető-e még.
A változás tétje nem technológiai – a hitelesség
egyre inkább rendszerszintű kérdéssé válik.**



MESTERSÉGESEN IGAZI

A digitális kultúra mind ez ideig azon alapult, hogy amit látunk és hallunk, valamilyen módon kapcsolatban áll a megtörtént eseményekkel. A generatív mesterséges intelligencia ezt az alapfeltevést kezdi átírni: ma már nem az a kérdés, hogy manipulálható-e a valóság, hanem az, hogy egyáltalán felismerhető-e még. A változás tétje nem technológiai. A hitelesség egyre inkább rendszerszintű kérdéssé válik.

A médiarendszer átalakulásának egyik legfontosabb hajtóereje ma a mesterséges intelligencia. Az MI nem különálló technológiai fejleményként jelenik meg, hanem a teljes médiakörnyezetet átszövő működési logikaként: a tartalom-előállításról a terjesztésen át egészen a befogadásig. Az automatizált szövegvírás, a képgenerálás és az ajánlórendszerek egyre nagyobb szerepet játszanak a tartalomgyártásban és -szerkesztésben. A generatív eszközök megjelenésével a tartalomkészítés lényegesen gyorsabbá és elérhetőbbé vált. Az MI segítségével szélesedett azok köre, akik mélyebb tudás, szakértelem nélkül is előállíthatnak tartalmakat. Ez a folyamat önmagában is aggályos lehet a valóság értelmezésének szempontjából, de emellett új kockázatokat is hoz magával. A deepfake-ek és más generált audiovizuális tartalmak nemcsak új médiatermékeket hoznak létre, hanem megkérdőjelezzik az eredeti tartalmak bizonyíthatóságának

fokát is. A jelenség hatására egyre több hazai szakmai fórum beszél a hitelesség válságáról. A Nemzeti Média- és Hírközlési Hatóság elemzése szerint a mesterségesen előállított tartalmak nemcsak a médiarendszert, hanem a társadalmi bizalmat is érintik: a felhasználók egyre nehezebben tudják eldönteni, mi tekinthető valóságnak.

A mindennapi online jelenlét során egyre nagyobb eséllyel találkozunk manipulált tartalmakkal, kétes eredetű üzenetekkel, közösségi médiában terjedő videókkal, dezinformációt tartalmazó hírekkel vagy akár álláspályázatokhoz csatolt bemutatkozó anyagokkal. A digitális tartalmak megbízhatósága többé nem magától értetődő. A helyzetet tovább nehezíti, hogy az ily módon generált tartalmak felismerése kifejezetten összetett feladat. Kutatások szerint az emberek csak korlátozott arányban képesek helyesen megítélni egy kép vagy videó hitelességét, ami arra utal, hogy vizuális ítéloképességünk egyre inkább elmarad a manipulációs technológiák fejlődésétől. Ebben a kontextusban az MI már nem pusztán eszköz, hanem strukturális tényező: átalakítja, mit tekintünk médiatartalomnak, és hogyan viszonyulunk hozzá.

TUDNI NEM ELÉG

Ebben a megváltozott helyzetben a médiatudatosság kulcsfontosságúvá válik, ám a felhasználók jelentős része nem ellenőrzi rendszeresen a tartalmak forrását vagy hitelességét,

és az információs túlterheltség és a gyors fogyasztási ritmus tovább csökkenti a kritikai feldolgozás esélyét.

A médiatudatosság viszont nem merülhet ki a digitális eszközök használatának képességében, sokkal inkább annak felismerését jelenti, hogy miként értelmezhető a rajtuk keresztül közvetített információk. Ide tartozik a forráskritika, a manipuláció felismerése, valamint annak megértése is, hogy egy adott tartalom milyen körülmények között jött létre. Ebben a helyzetben a technológiai megoldások, például a deepfake-detektálás vagy a forrásmegjelölés, önmagukban nem elegendők. Hatékonyságuk attól függ, hogy a felhasználók rendelkeznek-e azokkal az értelmezési készségekkel, fogódzókkal, amelyekkel ezek az eszközök egyáltalán használhatók.

A bizonytalanság így nem pusztán technológiai természetű. A változás mélyebb szinten a bizonyítás logikáját érinti: ha egy kép vagy hang már nem hordozza magában a hitelesség garanciáját, akkor új típusú kapaszkodókra van szükség. Ebben a kontextusban kerülnek ismét előtérbe azok a megoldások, amelyek a tartalmak eredetének azonosíthatóságát próbálják biztosítani. A deepfake-technológiák elterjedésével párhuzamosan megindult azoknak az eszközöknek a fejlesztése is, amelyek a manipulált tartalmak felismerését és a visszaélések visszaszorítását célozzák. Egyetemek, nagy technológiai vállalatok és startupok egyaránt dolgoz-

nak olyan algoritmusokon, amelyek képesek automatikusan azonosítani a mesterségesen előállított képeket és videókat. Ma már a felhasználók számára is elérhetőek különböző detektáló megoldások. Léteznek online felületek, ahol egy feltöltött videó elemzésével próbálják kimutatni a generált arcokat, valamint olyan böngészőbővítmények is, amelyek a megnyitott oldalak képi tartalmát vizsgálják, és figyelmeztetnek a gyanús elemekre.

A VALÓSÁG MÁSODLAGOS

A médiarendszer működésében meghatározó szerepet töltenek be a digitális platformok, amelyek egyre inkább az információhoz való hozzáférés elsődleges kapujává válnak. A tartalmak láthatóságát algoritmusok szabályozzák, amelyek nem a hitelességet, hanem a figyelmet és az aktivitást optimalizálják. Ez a működés kedvez az erős érzelmi reakciókat kiváltó, gyakran leegyszerűsített vagy torzított tartalmak terjedésének. A platformok így nem csupán közvetítik, hanem aktívan alakítják is a médiatartalmak szerkezetét és elérhetőségét. A hitelesség ebben a környezetben nem kizárólag tartalmi kérdés, hanem elosztási probléma is: az válik láthatóvá, ami az algoritmikus logikák szerint működik.

A platformok szerepe már nem korlátozódik a technikai infrastruktúra biztosítására, hanem a nyilvánosság működésének formálásában is meghatározóvá válik.



RÉGI ESZKÖZ, ÚJ TÉT: A VÍZJELEK VISSZATÉRÉSE

A digitális vízjel nem új jelenség: a képfeldolgozás és a médiatechnológia területén évtizedek óta alkalmazzák, elsősorban szerzői jogi és tartalomvédelmi célokra. A jelenlegi helyzet újdonsága nem maga a technológia, hanem annak megváltozott funkciója. A vízjel napjainkban már nemcsak azt jelzi, ki birtokolja a tartalmat, hanem azt, hogy milyen módon jött létre az adott tartalom. Ezért egyre nagyobb figyelem irányul az úgynevezett láthatatlan, vagyis a tartalomba beágyazott vízjelekre. Ezek nem jelennek meg a felhasználó számára érzékelhető formában, hanem a kép vagy hang statisztikai szerkezetében rejtve működnek. Magyar szakmai elemzések is rámutatnak arra, hogy az ilyen megoldások elsődleges célja a tartalom eredetének visszakövethetősége, különösen az MI által generált képek és videók esetében. A Szabályozott Tevékenységek Felügyeleti Hatósága szerint a deepfake-ek és más generatív tartalmak elleni fellépés egyik kulcseleme éppen az ilyen jelölési rendszerek fejlesztése lehet. Ebben az értelemben a vízjel a digitális bizalom egyik lehetséges technikai eszközévé válik. A kérdés viszont továbbra is nyitott: ha a jel nem látható, és technikai módon kerül a tartalomba, mennyiben tekinthető valódi kapaszkodónak a hitelesség megítélésében?

A digitális tartalmak megbízhatósága többé nem magától értetődő. Vizuális ítélőképességünk egyre inkább elmarad a manipulációs technológiák fejlődésétől.

A LÁTHATATLAN PECSÉT

A láthatatlan digitális vízjelek működése elsőre absztraktnak tűnhet, valójában azonban egyszerű elvre épül: a tartalom olyan módosítására, amely az emberi érzékelés számára észrevehetetlen, ugyanakkor algoritmikusan kimutatható. Ez a módosítás képek esetében a pixelek intenzitásának finom eltolásával, hangoknál az emberi fül számára nem hallható jelsorozat beágyazásával, szövegeknél pedig a tokenek (szóegységek) valószínűségi eloszlásának apró módosításával

történik. A generatív modellekhez kapcsolódó vízjelezés egyik legismertebb példája a Google DeepMind által fejlesztett SynthID, amely már a tartalom generálásának folyamatába építi be ezt a „rejtett mintázatot” – képek, hangok, videók és szöveges

tartalmak esetén egyaránt. A rendszer nem utólag jelöl, hanem a létrehozás során integrálja a jelölést: a tartalom eleve jelölten jön létre. Fontos ugyanakkor megjegyezni, hogy a SynthID nem általános MI-tartalom-detektor: elsősorban a saját maga által elhelyezett jeleket képes azonosítani, és megbízhatóan leginkább a Google saját modelljei által generált tartalmaknál működik.

Magyar informatikai és médiatechnológiai elemzések szerint az úgynevezett beágyazott jelölések jelenthetik az egyik leg-

ígéretesebb irányt az MI-tartalmak azonosításában, mivel nem igényelnek külön felhasználói beavatkozást, és jól automatizálhatók. Ugyanakkor ezek a rendszerek csak akkor működnek megbízhatóan, ha a generáló modellek beépítik a jeleket, és a platformok képesek azokat felismerni és értelmezni. Ha ez a lánc megszakad, a jel elveszíti funkcióját.

JEL VAN – DE KI OLVASSA?

A vízjelek működésének egyik kevésbé tárgyalt, mégis kulcsfontosságú aspektusa az értelmezés kérdése. Egy jel önmagában nem hordoz jelentést, ha nincs olyan rendszer, amely képes azt olvasni és értelmezni. A jelenlegi gyakorlatban a vízjelek értelmezése nagyrészt a platformok és fejlesztők kezében van. Ez azt eredményezi, hogy a hitelesség nem univerzális, hanem rendszerspecifikus: egy adott platform képes lehet azonosítani egy jelet, míg egy másik nem.

Ez a probléma a hazai szabályozási diskurzusban is megjelent. A Nemzeti Média- és Hírközlési Hatóság hangsúlyozza, hogy a technológiai megoldások önmagukban nem elegendők, ha nincs mögöttük egységes értelmezési keret és szabályozás. Másképp fogalmazva: a vízjel nemcsak technológiai kérdés, hanem intézményi is.

A generatív mesterséges intelligencia egyik legmélyebb következménye nem a tartalom minőségében, hanem a bizonyítás módjában érhető tetten. A korábbi médiakultúrában a

bizonyíték a tartalomhoz kötődött, egy felvétel önmagában hordozott bizonyító erőt. Ma ez a kapcsolat megbomlik. A tartalom egyre inkább elszakad az eredetétől, miközben a hitelesség külső rendszerekhez kapcsolódik, vízjelekhez, metaadatokhoz, ellenőrzési láncokhoz. A bizonyítás így többé nem a tartalomból, hanem az azt körülvevő rendszerekből következik. Ez alapvetően alakítja át a médiához való viszonyunkat. A „látom, tehát elhiszem” logikáját felváltja az ellenőrizhetőség logikája. Az ellenőrizhetőség viszont nem semleges. Attól függ, hogy kik építik és kik felügyelik azokat a rendszereket, amelyek a hitelességet garantálják.

KI GARANTÁLJA A VALÓSÁGOT?

A kérdés nem az, hogy megszűnik-e az igazság, hanem az, hogy miként válik felismerhetővé egy olyan közegben, ahol egyre inkább közvetített formában jelenik meg. A vízjelek, az eredetiigazolási rendszerek és a detektáló algoritmusok mind ugyanarra a problémára keresik a választ: hogyan lehet újraépíteni a bizalmat egy olyan médiatérben, ahol a tartalom leválik a létrejöttének körülményeiről? Ezek a technológiák azonban nem helyettesítik a bizalmat, hanem infrastruktúrát hoznak létre köré.

A hitelesség így fokozatosan technikai kérdéssé válik. Nem az lesz a döntő, hogy egy kép vagy hang önmagában hiteles-e, hanem az, hogy milyen rendszer hitelesíti. Platformok, fejlesztők és szabályozók együttműködése nélkül ezek a megoldások nem válhatnak egységes alapvetéssé, márpedig a bizalom csak akkor működik, ha közösen elfogadott. A magyar médiatudatossági diskurzus is ebbe az irányba mutat: a Nemzeti Média- és Hírközlési Hatóság hangsúlyozza, hogy a felhasználók kritikai kompetenciái legalább olyan fontosak, mint a technológiai megoldások. A hitelesség nem kizárólag a rendszerekben dől el, hanem abban is, ahogyan ezeket a rendszereket használjuk és értelmezzük.

A generatív korszak egyik paradoxona éppen ebben áll. Miközben egyre fejlettebb eszközöket alkotunk a valóság azonosítására, egyre kevésbé támaszkodhatunk magára a látványra. És ezzel együtt idővel a kérdés nem az lesz, hogy mi igaz, hanem az, hogy ki mondja meg, hogy az.

Szőke Andrea

Források: Nemzeti Média- és Hírközlési Hatóság: A deepfake pusztá létezése óvatosságra int és bizalmatlanságot szül. Médiapiaci Jelentés 2025. · Digital Hungary: A mesterséges intelligencia korában felértékelődik a médiatudatosság. · Kreatív: Felértékelődik a médiatudatosság a mesterséges intelligencia korában – megjelent az NMHH friss médiapiaci jelentése. · Expert Insights: The Top Deepfake Detection Solutions. · Spiralytics: Deepfake Statistics. · Nemzeti Média- és Hírközlési Hatóság: A deepfake-jelenséggel foglalkozik az NMHH új podcastsorozata.