

Ne hagyd magad becsapni: nyolc kritérium alapján be lehet azonosítani az AI-generált képeket

Ha egy kép illeszkedik a felhasználó meglévő hiedelmeihez vagy érzelmi tapasztalataihoz, könnyebben fogadja el azt hitelesnek. A mesterséges intelligencia által generált képek azonban beazonosíthatók gépi segítség nélkül, az árulkodó vizuális jelek értelmezésével is – állapítja meg a Corvinus Egyetem friss tanulmánya.

A Budapesti Corvinus Egyetem kutatója, Miskolczi Márk arra kereste a választ, milyen mechanizmusok miatt hisznek a közösségi média felhasználói olyan képeknek is, amelyeket valójában mesterséges intelligencia készített. A [tanulmány](#) a *Computers in Human Behavior* márciusi számában jelent meg.

Az eredmények szerint az AI-generált képek különösen akkor hatékonyak, amikor érzelmekre építenek, és egyszerre aktiválnak több kognitív torzítást, például a megerősítési torzítást, a lehorgonyozást, az ismerősségi hatást vagy a csoportgondolkodást. Ez erősebb és nehezebben kivédhető meggyőző hatást kelt, mint a szimpla félretájékoztatás. A felhasználók ugyanis gyakran mentális egyszerűsítésekre, úgynevezett kognitív heurisztikákra támaszkodnak, amikor gyorsan értelmeznek egy online képet. Ez a gyakorlat azonban növeli a manipuláció kockázatát, az érzelmi hatás pedig sok esetben felülírja a megtévesztett felhasználók kritikai gondolkodását, elaltatja éberségüket.

A kutató először 146 olyan képet azonosított, amelyek nagy valószínűséggel mesterséges intelligenciával készültek, és a kapcsolódó, több mint kilencezer Facebook-kommentet elemezte. A képeket speciális, intő vizuális jelek figyelembevételével szűrte ki – például szokatlan számú ujj, olvashatatlan szövegek, lebegő tárgyak vagy irreálisan sima textúrák alapján –, majd egy online AI-detektorral is ellenőrizte.

Így ismerd fel az AI-képeket

A képek AI-eredetére ez a nyolc vizuális kritérium utal, amely gépi algoritmus segítsége nélkül is ellenőrizhető:

1. **anatómiai** pontatlanságok: pl. szokatlan számú ujjak vagy fogak, természetellenes szögben álló végtagok;
2. aszimmetria és furcsaságok az **arcon**: pl. eltérő méretű vagy alakú szemek; a pupillák különböző irányba néznek; torz vagy hiányzó fül, természetellenes hajvonalak.
3. problémák a **tárgyaknál** és a környezetben: a tárgyat nem rendeltetésszerűen tartják; lebegő tárgyak; hiányzó vagy lehetetlen tükröződések, átfedések.
4. természetellenes **textúrák**: túl sima, természetellenesen tökéletes bőr; a szövet- és felületminták szabálytalanok, ismétlődőek.
5. **a fizika** és a gravitáció figyelmen kívül hagyása: tárgyak lebegnek; természetellenesnek ható folyadékok (pl. a pohárban lévő ital statikusnak vagy szilárdnak tűnik); az árnyékok iránya nem felel meg a fényforrás irányának.
6. **szövegekkel** és betűkkel kapcsolatos problémák: helytelen betűzés, torz betűk, értelmetlen karakterek; olvashatatlan vagy elmosódott szöveg.
7. a **szimmetria** eltűlése: irreálisan, túlzottan szimmetrikus arcok, tárgyak, természetellenes tökéletesség.

8. hibák a tömeg- és **csoportjelenetekben**: a háttérben látható emberek elmosódottak, torzak; az állatoknak anatómiai hibái vannak (pl. pluszlábak, hiányzó farok).

„A közösségi médiás AI-képek az érzelmi megtévesztés mesterei. A valódinak elhittetett AI-képek elterjedése hosszabb távon az online platformok iránti és a mesterséges intelligencia iránti általános bizalmat is gyengítheti. Emiatt sürgősen szabályozni kell a területet, például elengedhetetlen lenne kötelezően jelölni a mesterséges intelligencia által generált tartalmakat a közösségi médiában. Emellett minden generáció körében fejleszteni kell a digitális írástudást, ennek részeként érdemes megtanulni és alkalmazni a képek AI-eredetét kimutató nyolc kritériumot” – mondja Miskolczi Márk, a tanulmány szerzője.

Félrevezető nosztalgia, megnyugtató ismerősség

Az AI-képek leggyakoribb témái a nosztalgia és az együttérzést kiváltó történetek voltak. Gyakran jelentek meg idős emberekről, vidéki életről vagy családi kötelékekről szóló jelenetek. Népszerűek voltak a nehéz helyzetben lévő gyerekekről, elfeledett születésnapokról vagy különleges hobbiokról szóló képek. A vallási és spirituális motívumok szintén gyakran előfordultak. Például a nosztalgikus témák (évfordulók vagy vidéki életmód) a megerősítési torzítást használják ki pozitív asszociációk kiváltására, míg a lehorgonyzás felerősíti a kezdeti érzelmi hatást. A felhasználók többsége hitelesnek fogadta el ezeket a képeket. A reakciókban a legtöbbször empátia, nosztalgia vagy csodálat jelent meg. Sokan személyes történeteket osztottak meg vagy vigaszt próbáltak nyújtani a kép szereplőinek.

A tanulmány arra is rámutatott, hogy a mesterséges tartalmak körül gyakran automatizált fiókok jelennek meg. Ezek inspiráló üzenetekkel, vallási jókívánságokkal vagy születésnapi köszöntésekkel erősítik a posztok hitelességének látszatát, ami tovább növelheti a felhasználói aktivitást. Még azok között is, akik szkeptikusak is a képek hitelességét illetően, nagy arányban vannak olyanok, akik emberi hozzászólásként értelmezik a botok kommentjeit.

A kommentek AI-eredetére a következő kritériumok utalnak: a túl gyakori, túl gyors posztolás, szokatlan időpontokban, egyidejű azonos tevékenység több oldalon vagy csoportban, ismétlődő, sablonos vagy a kontextust figyelmen kívül hagyó megfogalmazás, a személyre szabott interakció hiánya, a kapcsolódó profilhoz tartozó információk túl általánosak, gyanúsak, a profilkép nem illik a profilhoz, és kevés barát, követő tartozik hozzá.

Sajtókapcsolat:

- Vajda Boglárka, szenior PR-szakértő
- Corvinus Kommunikáció
- +36 30 619 6633
- press@uni-corvinus.hu

© Budapesti Corvinus Egyetem
Példák AI generált képekre.



Eredeti tartalom: Budapesti Corvinus Egyetem

Továbbította: Helló Sajtó! Üzleti Sajtószolgálat

Ez a sajtóközlemény a következő linken érhető el:

<https://hellosajto.hu/?p=29742>