

Kako uočiti i razumjeti sadržaj generiran umjetnom inteligencijom ili digitalno izmijenjen sadržaj

Pristupačni i jednostavni modeli umjetne inteligencije (AI) mogu pomoći ljudima u učenju i stvaranju sadržaja, ali također mogu **povećati rizike koje dezinformacije predstavljaju** za otvorena društva i demokratski diskurs. Važno je spriječiti da naši zajednički informacijski prostori budu pretrpani dezinformacijama koje generira umjetna inteligencija i/ili koje su nastale digitalnim manipulacijama.

Dio rješenja je nova tehnologija kojom se može utvrditi porijeklo sadržaja i softveri za otkrivanje umjetno generiranog sadržaja. **Međutim, tehnološka rješenja daleko su od savršenih i potreban nam je rad neovisnih organizacija za provjeru činjenične točnosti kako bismo društvu pružili provjerene činjenice.**

Evo kratkog pregleda što profesionalne, neovisne organizacije za provjeru činjenične točnosti rade kako bi identificirale i razotkrile dezinformacije generirane umjetnom inteligencijom i što možete od njih naučiti.

AI generiran sadržaj bit će sve češći

Dezinformacije generirane umjetnom inteligencijom čine manji postotak svih tvrdnji koje istražuju profesionalne, neovisne organizacije za provjeru činjenične točnosti. Organizacije za provjeru činjenične točnosti češće se u svome radu susreću s digitalno izmijenjenim sadržajem.

Međutim, u internoj anketi članica EFCSN-a, **većina organizacija za provjeru činjenične točnosti složila se da će sadržaj generiran umjetnom inteligencijom i digitalno izmijenjeni sadržaj samo rasti** i biti sve relevantniji u budućnosti. Nedavni primjeri u kontekstu europskih izbora podupiru ova predviđanja.

OBJAŠNJENJE: *Digitalno izmijenjeno* se odnosi na bilo koji oblik sadržaja koji je značajno izmijenjen kako bi se manipuliralo ili promijenilo poruku koju je izvorno prenosio, uključujući uređivanje sadržaja AI alatima. To ne uključuje izmjene radi jasnoće ili kvalitete.

Generirano umjetnom inteligencijom se odnosi na bilo koji oblik sadržaja koji je stvorio sustav umjetne inteligencije.



Tehnologija se brzo razvija, ali ne možemo se osloniti samo na nju.

Stručnjaci za umjetnu inteligenciju i profesionalne organizacije za provjeru činjenične točnosti se slažu: **Alati za otkrivanje AI sadržaja sami po sebi nisu dovoljni za utvrđivanje ili razotkrivanje sadržaja generiranog umjetnom inteligencijom ili digitalno izmijenjenog sadržaja.**

Prije uporabe tih alata, stručnjaci preporučuju upoznavanje s alatima za generiranje sadržaja umjetnom inteligencijom i alatima za otkrivanje takvog sadržaja. Uz razumijevanje procesa učenja AI modela kao i poznavanja statistike, organizacije za provjeru činjenične točnosti mogu prepoznavati prednosti i slabosti alata i vjerojatnost uspješne provjere. **Alati za otkrivanje mogu biti korisna polazna točka.**

Inicijative za utvrđivanje **porijekla sadržaja** kao što su C2PA specifikacije mogu pomoći u potvrđivanju izvora i povijesti medijskog sadržaja, ali stavljanje vodenog žiga i verifikacija nisu neoborivi dokazi autentičnosti sadržaja.

Kako dezinformacije koje generira umjetna inteligencija utječu na ljude

„Svaki put kad reagirate svojom intuicijom, zaobilazite proces promišljanja.“

– Christine Dugoin*

PSIHOLOGIJA: Operacije utjecaja često su osmišljene tako da iskoriste pristranosti utemeljene u psihologiji.

Razumijevanje vlastitih pristranosti i onih svoje publike može pomoći u suzbijanju dezinformacija.

CILJEVI: Zašto bi se zlonamjerni akteri mogli oslanjati na umjetnu inteligenciju za stvaranje ili širenje dezinformacija? Kakav utjecaj žele postići u stvarnom svijetu?

- Proširiti svoj doseg u drugu zemlju ili zajednicu?
- Izbjeći otkrivanje ili preplaviti organizacije za provjeru činjenične točnosti generiranjem brojnih varijanti sličnih tvrdnji?
- Utjecati na javno mnijenje putem mreža neautentičnih računa?

* Christine Dugoin istraživačica je informacijskog utjecaja na La Sorbonneu.

Razotkrivanje dezinformacija zahtijeva sveobuhvatan pristup i razumijevanje specifičnosti

Dakle, ako alati za otkrivanje ne funkcioniraju, što funkcionira? Važno je razumjeti kontekst određene tvrdnje, jednako kao i njen sadržaj. Profesionalne organizacije za provjeru činjenične točnosti imaju stručnjake s potrebnim istraživačkim vještinama. Evo nekoliko savjeta.

„Alati za otkrivanje nikada neće funkcionirati za 100 % sadržaja – ne očekujem da će ikada funkcionirati.“
– Henk van Ess**



UZMITE U OBZIR IZVOR: Možete li potvrditi njihov identitet? O čemu razgovaraju i kakav sadržaj dijele? Tko je u interakciji s njihovim sadržajem? Kakav bi učinak ovaj sadržaj mogao imati na čitatelje?



UTVRDITE VJERODOSTOJNOST: Neovisna provjera informacija s vjerodostojnim izvorima kao što su stručnjaci s praktičnim iskustvom u određenom području. Ima li ono što je prikazano smisla na temelju vašeg znanja?



Upotrijebite **FORENZIČKE MEDIJSKE** tehnike kao dopunu tradicionalnom istraživanju, kao što su: rudarenje podataka, geolociranje, biometrijsko prepoznavanje, analizu uzoraka i drugo.



UČITE I PRILAGOĐAVAJTE SE: Kreatori netočnih informacija i dezinformacija generiranih umjetnom inteligencijom stalno se prilagođavaju. Prilagodite svoj pristup promjenjivom krajoliku.

PODIJELITE SVOJ RAD

Uz razotkrivenu tvrdnju, stručnjaci preporučuju transparentan prikaz analize i pružanje poveznica na izvore. To može pomoći čitateljima da prate i nastave istragu te razumiju nijanse dezinformacijskih narativa. U nekim je slučajevima istraga važnija od toga je li sadržaj napisan umjetnom inteligencijom.

** Henk van Ess stručnjak je za OSINT i tehnike provjere činjenica.

Kratki vodič: Što činiti, što ne činiti i tragovi

Sljedeći bi tragovi mogli ukazivati na to da je dio sadržaja generiran umjetnom inteligencijom ili digitalno izmijenjen. Zajedno s drugim savjetima spomenutim u ovom vodiču (kontekst, tehnike istraživanja i alati za otkrivanje), mogu vam pomoći da shvatite istinu iza onoga što vidite.

Tekst

- Često (ali ne uvijek) ima **bolju gramatiku** nego kada čovjek piše.
- Vjerojatno će se koristiti **pretjerano formalnim ili strukturiranim jezikom**, posebno za kontekst društvenih medija.
- **Prekomjerni prilozi ili pridjevi**.
- Nedostatak ljudskih emocija, humora, sarkazma i idiomatskih izraza.
- Možda **nedostaju određeni detalji** (imena, datumi, lokacije) ili originalne ideje.
- Najvažnije: jesu li činjenične tvrdnje navedene u tekstu točne?

Videozapis

- Nemojte upotrebljavati detektor slika koje generira umjetna inteligencija na fotografijama iz videozapisa.
- Pogledajte **izraze lica i pokrete**, kao što je treptanje, te odgovara li kretanje usta zvuku.
- Može biti obilježeno **oštrim prijelazima ili rezovima**.

Zvuk

- **Usporedite sumnjivi zvuk s autentičnim uzorkom** s pomoću alata koji mogu otkriti razlike u govoru i uzorcima disanja, intonaciji...
- Kada upotrebljavate detektore, izbjegavajte uzorke zvuka loše kvalitete sa statičkim ili pozadinskim šumom.
- Može biti obilježen **neprirodnim ili mehaničkim govornim obrascima**, nedostatkom stanki ili prirodnog disanja.

Slika

- Potražite područja s **neprirodnim detaljima**: savršena koža, zamagljena pozadina, neprirodna ljepota ili svjetlo i neobičnosti kao što su dodatni prsti.
- **Potražite vodeni žig** iz uobičajenih generatora slika.
- Obratite pažnju na detalje onoga što je prikazano: je li sve logično? Je li sve pripadajuće?
- Kada upotrebljavate detektore, **odlučite se za verziju slike visoke razlučivosti ili koja je ranije učitana, a ne za onu koja je već puno puta podijeljena**.