

# Kako uočiti i razumjeti sadržaj generisan umjetnom inteligencijom ili digitalno izmijenjen sadržaj

Pristupačni i laki za korištenje, AI modeli mogu biti korisni za učenje i kreiranje sadržaja, ali mogu i **povećati rizike koje dezinformacije predstavljaju** za otvorena društva i demokratski diskurs. Potrebno je spriječiti da naš zajednički informacijski prostor bude zatrpan dezinformacijama koje su generisane umjetnom inteligencijom i/ili koje su nastale digitalnim manipulacijama.

Dio rješenja su nove tehnologije kojim se može utvrditi porijeklo sadržaja i softveri za otkrivanje umjetno generisanog sadržaja. **Međutim, tehnološka rješenja su daleko od savršenih i potrebne su nam nezavisne organizacije za provjeru činjenica koje će društvu ponuditi provjerene činjenice.**

Napravili smo kratak pregled onoga što nezavisne organizacije za provjeru činjenica rade kako bi identifikovale i raskrinkale dezinformacije generisane umjetnom inteligencijom, kao i onoga što od njih možete naučiti.

## Sadržaj generisan umjetnom inteligencijom bit će sve češći

Dezinformacije generisane umjetnom inteligencijom trenutno predstavljaju manji dio svih tvrdnji koje istražuju nezavisne organizacije za provjeru činjenica. Digitalno izmijenjeni sadržaj je daleko više zastupljen u njihovom radu.

Ali: u internoj anketi članova EFCSN-a, **većina organizacija za provjeru činjenica se složila da će sadržaj generisan umjetnom inteligencijom i digitalno izmijenjeni sadržaj postati sve relevantniji** u budućnosti. [Nedavni primjeri](#) u kontekstu evropskih izbora potvrđuju ovakva predviđanja.

**Objašnjenje:** *Digitalno izmijenjeno* se odnosi na bilo koji oblik sadržaja koji je značajno izmijenjen kako bi se izmanipulisala ili promijenila poruka koju je taj sadržaj izvorno prenio, uključujući i izmjene pomoću AI alata. Ovo ne uključuje izmjene u cilju bolje jasnoće ili kvaliteta.

*Generisano umjetnom inteligencijom* se odnosi na bilo koji oblik sadržaja koji je kreiran pomoću sistema vještačke inteligencije.



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

Stručnjaci za umjetnu inteligenciju i organizacije za provjeru činjenična su saglasni: **alati za detekciju sami po sebi nisu dovoljni za razotkrivanje sadržaja generisanog umjetnom inteligencijom ili digitalno izmijenjenog sadržaja.**

Prije korištenja ovih alata, stručnjaci/kinje preporučuju da se prvo upoznamo sa alatima za generisanje sadržaja umjetne inteligencije i alatima za otkrivanje takvog sadržaja sadržaja. Uz razumijevanje načina na koji se AI modeli obučavaju, kao i malo statistike, organizacije za provjeru činjenica mogu prepoznati prednosti i nedostatke alata te vjerovatnoću njihovog uspjeha. **Alati mogu biti korisna polazna tačka.**

**Inicijative za utvrđivanje porijekla sadržaja** kao što je C2PA Specifications, mogu pomoći u potvrđivanju izvora i istorije medijskog sadržaja, ali vodeni žig i verifikacija nisu nepobitni dokazi.

## Kako dezinformacije generisane umjetnom inteligencijom utiču na ljude

*“Svaka vaša reakcija koja, ako mogu tako reći, dolazi iz unutrašnjeg osjećaja, izmiče vašoj sopstvenoj refleksiji.”*

- Christine Dugoin\*

**PSIHOLOGIJA:** Operacije uticaja su često osmišljene na način da iskorištavaju pristranosti utemeljene u psihologiji.

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

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

- Proširiti doseg na drugu zemlju ili zajednicu?
- Izbjeći otkrivanje ili preopteretiti organizacije za provjeri činjenica generisanjem različitih verzija sličnih tvrdnji?
- Uticati na mišljenja ili uvjerenja, uspostavljanjem kredibiliteta putem mreža neautentičnih naloga?

\* Christine Dugoin je istraživačica informacijskog uticaja na La Sorbonne.

# Provjera činjenica zahtijeva višedimenzionalni pristup i razumijevanje specifičnosti

Dakle, ako alati za detekciju ne funkcionišu, šta funkcioniše? Razumijevanje konteksta tvrdnje je jednako važno koliko i razumijevanje njenog sadržaja. Profesionalne organizacije za provjeru činjenica imaju stručnjake s potrebnim istraživačkim vještinama. Evo nekoliko savjeta

*“Alati za detekciju nikad ne funkcionišu 100% – ne očekujem da ikad i hoće.”*

– Henk van Ess\*\*



**UZMITE U OBZIR IZVOR:** Možete li potvrditi njihov identitet? O čemu pričaju i šta dijele? Ko ostvaruje interakciju sa njihovim sadržajem? Kakav efekat ovaj sadržaj može imati na čitatelje/ke?



**UTVRDITE KREDIBILITET:** Izvršite nezavisnu provjeru tačnosti informacija kod kredibilnih izvora kao što su eksperti/ce sa praktičnim iskustvom iz date oblasti. Na osnovu vašeg znanja, da li prikazano ima smisla?



Koristite **FORENZIČKE MEDIJSKE** tehnike kao dopunu tradicionalnom istraživačkom novinarstvu i dokumentarnom istraživanju. Neke od ovih tehnika uključuju: *scraping* podataka, geolociranje, biometrijsko prepoznavanje, analizu uzoraka i drugo.



**UČITE I PRILAGODITE SE:** Kreatori dezinformacija generisanih umjetnom inteligencijom se konstantno prilagođavaju. Prilagodite i vi svoj pristup ovim promjenama.

## PODIJELITE SVOJ RAD

Eksperti/ce preporučuju da, uz opovrgnutu tvrdnju, podijelite i transparentnu analizu i linkove na izvore. Na taj način će čitatelji/ke moći pratiti tok istrage i razumjeti slojevitost narativa.

U nekim slučajevima, istraga je važnija od činjenice da je sadržaj napisan uz pomoć vještačke inteligencije.

\*\* Henk van Ess je ekspert za OSINT i tehnike provjera činjenica.

# Šta treba, šta ne treba i koji su znakovi

Ovi znakovi mogu ukazivati na to da je sadržaj sadržaja generisan umjetnom inteligencijom ili digitalno izmijenjen. Zajedno sa ostalim savjetima iz ovog vodiča (kontekst, istraživačke tehnike i alati za detekciju), oni vam mogu pomoći da shvatite istinu iza onoga šta vidite.

## Tekst

- Često (ne i uvijek) je **gramatički ispravniji** od onoga kojeg je napisao čovjek.
- Vjerovatno će se koristiti **pretjerano formalan ili struktuiran jezik**, posebno u kontekstu društvenih medija.
- **Pretjerano korištenje priloga ili pridjeva.**
- Nedostatak ljudskih emocija, humora, sarkazma ili idiomatskih izraza.
- Mogući **nedostatak preciznih detalja** (imena, datuma, lokacija) ili originalne ideje.
- Najvažnije: da li su informacije navedene u tekstu tačne?

## Video

- Nemojte koristiti detektor slika koje generira umjetna inteligencija na fotografijama iz videozapisa.
- Pogledajte **izraze lica i pokrete**, kao što je npr. treptanje te da li su pokreti usana usklađeni sa zvukom.
- Može sadržavati **oštre prelaze ili rezove.**

## Audio

- **Uporedite sumnjivi audio zapis sa autentičnim uzorkom** uz pomoć alata koji mogu detektovati razlike u načinu govora i disanja, intonaciji...
- Kada koristite detektore, izbjegavajte audio uzorke niske kvalitete sa statičkim ili pozadinskim šumom.
- Može sadržavati **neprirodan ili mehanički govor**, bez prirodnih pauza u govoru ili pauza za disanje.

## Slike

- Uočavaju li se dijelovi sa **neprirodnim detaljima**: savršena koža, zamagljena pozadina, neprirodna ljepota ili svjetlo, neobičnosti poput dodatnih prstiju.
- **Potražite vodeni žig** nekih od poznatih generatora slika.
- Obratite pažnju na detalje prikazanog: da li je logično? Da li je prikladno?
- Ako koristite detektore, **odaberite sliku visoke rezolucije ili prvobitno objavljenu verziju, umjesto one koja je bila podijeljena jednom ili više puta.**