

# Kako uočiti i razumeti sadržaj kreiran uz pomoć veštačke inteligencije ili digitalno izmenjen sadržaj

Pristupačni i jednostavni za upotrebu modeli veštačke inteligencije pomažu ljudima da uče i kreiraju sadržaj, ali takođe **povećavaju rizike koje dezinformacije predstavljaju** za demokratska društva. Važno je sprečiti da zajednički informacioni prostori budu preplavljeni dezinformacijama koje su kreirane uz upotrebu veštačke inteligencije ili su digitalno izmenjene.

Deo rešenja su nove tehnologije uz pomoć kojih je moguće utvrditi poreklo sadržaja i digitalni alati za detekciju. Međutim, **tehnološka rešenja nisu savršena i potrebno je učešće nezavisnih i profesionalnih faktčekera koji društvu ukazuju na proverene i istinite informacije.**

Evo kratkog pregleda onoga što nezavisni faktčekeri rade kako bi identifikovali i razotkrili dezinformacije kreirane uz pomoć veštačke inteligencije i onoga šta možete naučiti od njih.

## Sve je više sadržaja kreiranog uz pomoć veštačke inteligencije

Trenutno, dezinformacije kreirane uz upotrebu veštačke inteligencije čine mali procenat od ukupnog broja tvrdnji koje faktčekeri analiziraju, dok je digitalno izmenjen sadržaj prisutniji u njihovim proverama.

Međutim, u internom istraživanju među članovima EFCSN-a, većina faktčekera se složila da će sadržaj kreiran uz pomoć veštačke inteligencije i digitalno izmenjeni sadržaj postati sve prisutniji u budućnosti. To potvrđuju i nedavni primeri dezinformacija koje se dele u kontekstu evropskih izbora.

**NAUČI:** *Digitalno izmenjen* sadržaj odnosi se na bilo koji oblik sadržaja koji je značajno izmenjen kako bi se manipulisala ili promenila poruka koju je prvobitno prenosio, uključujući izmene pomoću alata veštačke inteligencije. Ovo ne uključuje izmene radi jasnoće ili kvaliteta.

*Sadržaj kreiran uz pomoć veštačke inteligencije* odnosi se na bilo koji oblik sadržaja koji je kreirao sistem veštačke inteligencije.



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

Stručnjaci za veštačku inteligenciju i fektčekerri se slažu da **alati za detekciju veštačke inteligencije nisu dovoljni kako bi se razotkrio sadržaj koji je kreirala veštačka inteligencija ili digitalno izmenjeni sadržaj.**

Pre upotrebe ovih alata, eskperti savetuju da se bolje upoznate sa generatorima i detektorima sadržaja baziranim na veštačkoj inteligenciji. Uz razumevanje o tome kako se modeli treniraju i uz osnovno poznavanje statistike, fektčekerri mogu da prepoznaju prednosti i nedostatke alata, kao i verovatnoću njegovog uspeha. U svakom slučaju, **alati mogu biti koristan početak.**

Inicijative za **poreklo sadržaja**, poput C2PA specifikacija, mogu pomoći u potvrđivanju izvora i istorije sadržaja, ali vodeni žigovi i verifikacija nisu nepogrešivi.

## Kakav uticaj imaju dezinformacije koje je kreirala veštačka inteligencija?

*“Svaki put kada reagujete instiktivno, ako smem da kažem, reagujte bez dovoljno promišljanja.”*

– Christine Dugoin\*

**PSIHOLOGIJA:** Operacije za postizanje uticaja često su osmišljene da iskoriste psihološke pristrasnosti.

Razumevanje vaših, kao i pristrasnosti vaše publike, može pomoći u suzbijanju dezinformacija.

**CILJEVI:** Zašto se loši akteri oslanjaju na veštačku inteligenciju da kreiraju ili šire dezinformacije? Koji uticaj nameravaju da postignu?

- Žele da prošire doseg do druge države ili zajednice?
- Da li izbegavaju detekciju ili preplavljaju fektčekere generisanjem mnogo varijanti sličnih tvrdnji?
- Utiču li na mišljenja ili uverenja tako što uspostavljaju kredibilitet putem mreža lažnih naloga?

\* Christine Dugoin je istraživačica u oblasti informacionog uticaja na Sorboni.

# Razotkrivanje zahteva višedimenzionalni pristup i razumevanje nijansi

Dakle, ako alati za detekciju ne funkcionišu, šta onda? Treba razumeti kontekst tvrdnje koliko i njen sadržaj. Fektčakeri su eskperti za istraživanja i imaju neophodne veštine. Evo nekoliko saveta.

*“Alati za detekciju nikada neće raditi 100% tačno, i ja to ne očekujem”*  
– Henk van Ess\*\*



**RAZMISLITE O IZVORU:** Možete li utvrditi identitet izvora? O čemu priča i šta deli? Ko ima interakciju sa njegovim sadržajem? Kakav bi uticaj ovaj sadržaj mogao imati na publiku?



**UTVRDITE KREDIBILITET:** Proverite informacije i kod verodostojnih izvora poput stručnjaka sa praktičnim iskustvom u datoj oblasti. Da li ono što je prikazano ima smisla na osnovu vašeg znanja?



Koristite **FORENZIČKE** tehnike kako biste dopunili tradicionalno istraživačko novinarstvo i dokumentaciona istraživanja. Neke od tehnika su skidanje podataka, geolokacija, biometrijsko prepoznavanje, analizu obrazaca i još mnogo toga.



**UČI I PRILAGOĐAVAJ SE:** Kreatori dezinformacija uz pomoć veštačke inteligencije se stalno prilagođavaju. Prilagodite i vi svoj pristup situaciji koja se stalno menja.

## DELITE SVOJ RAD

Pored razotkrivene tvrdnje, stručnjaci preporučuju pružanje transparentne analize i linkova ka izvorima. Ovo može pomoći čitaocima da prate istragu i razumeju sve slojeve priče. U nekim slučajevima, istraživanje je važnije nego da li je sadržaj napisala veštačka inteligencije.

\*\* Henk van Ess je stručnjak za OSINT (otvorene izvore informacija na internetu) i tehnike proveravanja činjenica.

Sledeći tragovi mogu ukazati na to da je određeni sadržaj generisan uz pomoć veštačke inteligencije ili digitalno izmenjen. Uz ostale savete navedene u vodiču (kontekst, istraživačke tehnike i alate za detekciju), mogu vam pomoći da razumete istinu iza onoga što vidite.

## Tekst

- Često (ali ne uvek) ima **bolju gramatiku** nego čovek.
- Uglavnom koristi **preterano formalan jezik**, posebno u kontekstu društvenih mreža.
- **Preterana upotreba priloga ili prideva.**
- Odsustvo ljudske emocije, humora, sarkazma i fraza.
- Mogu nedostajati specifični detalji (imena, datumi, lokacije) ili originalne ideje.
- Najvažnije: da li su činjenice navedene u tekstu tačne?

## Video

- Nemojte koristiti detektor slika generisanih veštačkom inteligencijom na fotografijama iz video zapisa.
- Posmatrajte izraze lica i pokrete, poput treptaja, kao i da li pokreti usta odgovaraju audio zapisu.
- Može sadržati oštre prelaze ili rezove.

## Audio

- **Uporedite sumnjivi audio sa autentičnim uzorkom** koristeći alate koji mogu otkriti razlike u govoru i disanju, intonaciji...
- Prilikom korišćenja detektora, izbegavajte uzorke niskog kvaliteta sa pozadinskom bukom.
- Mogu ga karakterisati **neprirodni ili mehanički govor**, nedostatak pauza ili prirodnog disanja.

## Slike

- Tražite delove sa **neprirodnim detaljima**: savršena koža, zamućene pozadine, neobična lepota ili osvetljenje, kao i neobičnosti poput dodatnih prstiju.
- **Poražite vodeni žig poznatih generatora slika.**
- Obratite pažnju na detalje onoga što je prikazano: da li je logično? Da li je prikladno?
- Prilikom upotrebe detektora, birajte verziju slike visoke rezolucije ili verziju koja je postavljena ranije u odnosu na onu koja se masovno deli.