



AI stvara i širi lažne vijesti, ali hoće li ih uskoro moći i prepoznati?

U razgovoru s domaćim stručnjacima otkrivamo što lažnu vijest čini uspješnom, mogu li AI alati spriječiti njihov nastanak i što je Big Tech poduzeo da se riješi jedan od najvećih problema današnjice.

Istražili smo primjene [AI-ja u političkim kampanjama](#) i koliku opasnost danas predstavlja sve [naprednija deepfake tehnologija](#), a **da nas pritom malo tko štiti od nje**. Što se pak događa s lažnim vijestima? Znamo da je [ova godina raj za njih](#) zbog mnoštva izbora, a Hrvatska čak nešto i pokušava po tome pitanju. **Trenutno razvijamo dva alata za prepoznavanje lažnih vijesti u Zagrebu i Osijeku**, a to je tek dio aktivnosti.

Kako bismo saznali više o korištenju **umjetne inteligencije u svrhu širenja i stvaranja lažnih vijesti**, razgovarali smo s **Petrom Vidovom** iz Faktografa, prof. dr. sc. **Josipom Jobom** s Fakulteta elektrotehnike, računarstva i informacijskih tehnologija u Osijeku, i **Ivom Šalamon**, voditeljicom tima za *fact-checking* Hrvatske udruge digitalnih izdavača.

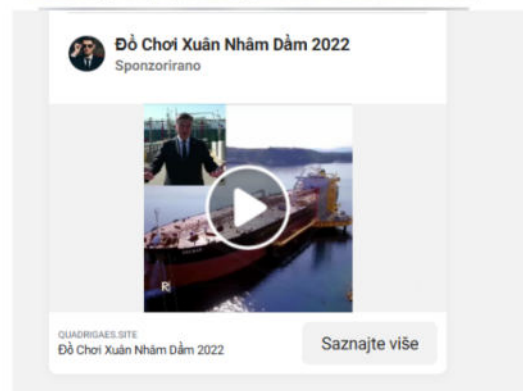
Iza dezinformacija ne stoje alati, već sposobnost izvrtanja istine

Novinari Faktografa izvještavali su o **velikom broju dezinformacija kreiranih pomoću umjetne inteligencije**. Samo neki od zanimljivih primjera su:

- premijer **Andrej Plenković** pozivao je na ulaganje u tvrtku Prvo plinarsko društvo jer tako svi mogu s uložених 250 eura zaraditi 1500 eura tjedno,
- supruga ukrajinskog predsjednika **Olena Zelenska** je novcima poreznih obveznika kupila Bugatti Tourbillon kojeg je predstavio **Mate Rimac**,

- francuski predsjednik **Emmanuel Macron** ljubi muškarca na jahti.

A mnoge smo primjetili i mi pa smo tako mogli vidjeti “[velike skandale](#)” [Vladimira Kočiša Zeca](#), a [Mojmira Pastorčić](#) i bivša predsjednica [Kolinda Grabar Kitarović](#) [pozivale su na investiciju od 250 eura u “kripto platformu”](#).



Uglavnom se radilo o slučajevima u kojima su se lik i glas sintetički generiranih imitacija hrvatskih političara (i drugih javnih osoba) koristili u svrhu pokušaja prijevare, tj. krađe novca ili osobnih podataka. Na Faktografu su uočili AI dezinformacije koje imaju politički kontekst, ali takav sadržaj se **redovito odnosi na inozemnu politiku i uglavnom se preuzima od stranih korisnika društvenih mreža**, objašnjava Petar.



Petar Vidov

Faktograf, glavni urednik Klimatskog portala

Deepfake ostaje najpopularniji oblik AI dezinformacija

Najraširenije digitalne dezinformacije još uvijek su najčešće vrlo jednostavne, nastale bez korištenja AI alata. Uspješnu dezinformaciju ne čini tehnološka sofisticiranost sadržaja, **već sposobnost izvrtanja istine** (najraširenije dezinformacije uglavnom su djelomično utemeljenje u činjeničnoj stvarnosti) i **manipuliranja emotivnim reakcijama** javnosti na internetu, ističe Petar:

Sadržaj koji je emotivno nabijen, koji je napisan s ciljem da kod ljudi potakne snažnu emotivnu reakciju (npr. da ih naljuti), brzo se i lako širi internetom.

S druge strane, **najpoznatija primjena umjetne inteligencije u stvaranju lažnih vijesti je [deepfake](#)**. Dok je danas takve materijale još i moguće razlikovati od stvarnosti, u

budućnosti će to širokim masama biti puno teže bez alata koji će obavljati tu ulogu – prepoznavati lažne slike i tekstove, odnosno lažne vijesti, ističe Josip.

Može li AI pomoći u traženju AI dezinformacija?

Koliko je Petru poznato, za sada ne postoje **AI alati koji su sposobni samostalno prepoznavati dezinformacije**. Različite organizacije diljem svijeta eksperimentiraju s velikim jezičnim modelima, ali fokus je na razvoju alata koji mogu pomoći *fact-checkerima* i istraživačima u obavljanju njihovog posla (poput npr. alata za transkripciju zvuka u tekst ili za brže i lakše filtriranje velike količine sadržaja).

Ipak, na jednom takvom alatu radi se u Osijeku u suradnji s HUDI-jem. Projektni tim FACT@FERIT, koji vodi Josip, nastoji razviti alat koji će **prepoznavati činjenice u vijestima iz područja elektrotehnike, računarstva i informacijske tehnologije** čija će se istinitost moći provjeriti kroz metodologiju definiranu u okviru projekta:

U prepoznavanju činjenica iz vijesti koristit će se umjetna inteligencija kako bi se ubrzala obrada i povećali kapaciteti sustava.



prof. dr. sc. Josip Job

Redoviti profesor na FERIT-u

Inače, HUDI radi na još jednom *fact-checking* projektu u suradnji s Agencijom za elektroničke medije, Ministarstvom kulture i medija i Sveučilištom VERN' – TurFacto. Projekt je usmjeren na jačanje vjerodostojnosti informacija u turizmu.

Big Tech mora spriječiti širenje dezinformacija, ali trenutno ne radi dovoljno

Sve više tehnoloških organizacija najavljuje da će AI sadržaj generiran uz pomoć njihovih alata biti označen *watermarkom*, **kako bi korisnici mogli prepoznati da se ne radi u autentičnim foto, video ili audio zapisima**, navodi Petar:

Ako se te najave ostvare, takvo bi tehnološko rješenje moglo biti korisno.

Naravno, neće biti dobro ako ostane samo na tome. AI je nova, **potencijalno vrlo disruptivna tehnologija pa bi razvoj takvih proizvoda trebala pratiti i adekvatna regulacija**, smatra Petar. Samo neke od tih regulacija su europski [DSA](#) i [AI akt](#), a SAD trenutno radi na vlastitoj verziji [zakona o umjetnoj inteligenciji](#).

Meta radi osjetno više nego Google, ali, prema Petru, nijedna ne čini dovoljno. Isto se odnosi i za sve druge velike platforme čija se infrastruktura koristi za širenje svakojakog “smeća” – od govora mržnje preko toksične propagande do dezinformacijskih kampanja:

Neki ulažu više resursa u suzbijanje štetnih sadržaja, neki manje, ali očigledna raširenost ne samo dezinformacija, već i različitih internetskih prijevара te sličnog sadržaja, jasno pokazuje da to što rade nije dovoljno.

S druge strane, **Iva smatra kako su obje kompanije poduzele su značajne korake kako bi smanjile širenje dezinformacija** na svojim platformama, ali postoje i kritike koje ukazuju na to da bi mogli učiniti više. Za sada ostajemo na ovim primjerima:

- **Meta je razvila Third-Party Fact-Checking Program**, u kojem su uspostavili suradnju s nezavisnim *fact-checkerima* kako bi označavali i smanjivali doseg lažnih vijesti na svojim platformama (Facebook, Instagram, WhatsApp), a kada *fact-checkeri* označe sadržaj kao lažan, Meta smanjuje njegovu vidljivost i korisnicima prikazuje upozorenja.
- **Google podržava brojne fact-checking organizacije** kroz različite grantove i inicijative. Također, kroz Google News, koristi oznake za provjerene članke kako bi korisnici mogli lakše prepoznati vjerodostojne izvore.

Dok su algoritmi Mete i Googlea postali sofisticiraniji, a suradnja s *fact-checking* organizacijama intenzivnija, volumen i složenost problema ukazuju **na to da bi moglo biti prostora za dodatna poboljšanja**, smatra Iva.



Iva Šalamon

Voditeljica HUDI-jeva tima
za fact-checking

Što Faktograf preporučuje građanima?

Najbolje što prosječni građanin može napraviti da se zaštiti jest da **kritički propituje sadržaj s kojim dolazi u kontakt** i postavi sebi ova pitanja:

- tko je izvor informacije?
- kakva je kredibilnost tog izvora?
- je li informaciju moguće potvrditi kod drugih izvora?

Na kraju Petar poziva sve one koji su u dvojbi **da se obrate profesionalnoj fact-checkerskoj organizaciji poput Faktografa**. Mnoge teme na kojima su radili nastale su na temelju upita čitatelja.

Članak je dio serijala [“Umjetna inteligencija i politika”](#) koji je realiziran uz podršku Agencije za elektroničke medije kroz projekt “Poticanja novinarske izvrsnosti”.