



Prevara poznatog lica

AI generirani sadržaj u SEE regiji

Sadržaj

Uvod	3
AI generirani sadržaj u SEE regiji	12
Hrvatska	14
Bosna i Hercegovina	18
Crna Gora	20
Srbija	24
Narativi o AI-ju u SEE regiji	29
Zaključak	38

Analizu pripremila:

Tajana Broz

Urednik:

Petar Vidov

Prijelom i dizajn:

Jordi Ilić

Izdavač:

Faktograf – udruga za informiranu javnost
Zagreb, rujan 2025. godine

Uvod

AI je svuda oko nas, a istovremeno je vrlo teško objasniti o čemu zapravo govorimo kad govorimo o AI-ju.

Sam akronim dolazi od engleskog termina *Artificial Intelligence*, tj. umjetna inteligencija. Danas se termin AI uglavnom koristi za opisivanje velikih jezičnih modela (*large language model*, LLM), softverskih proizvoda koji su sposobni na zahtjev generirati tekst, sliku, zvuk ili video.

U pitanju su, kako mnogi [stručnjaci vole istaknuti](#), strojevi kojima nikakva inteligencija nije imanentna, već se bave statističkom analizom. Temeljem sadržaja na kojem su trenirani i upita koji im je postavljen predviđaju što korisnik želi i to mu pokušavaju isporučiti. Uočavaju obrasce obrađujući velike količine podataka i koriste spoznaje o tim obrascima da bi generirali sadržaj.

Drugim riječima, ono što danas nazivamo AI-jem nije stvarna umjetna inteligencija, ali ima potencijal da to postane. Usljed takve terminološke konfuzije, ne iznenađuje činjenica da sam pojam “umjetne inteligencije” nije još do kraja definiran. Većina postojećih definicija uključuje “sposobnost stroja da oponaša ljudsko ponašanje”, odnosno da učinkovito i neovisno rješava složene probleme.

Iako stvarnom umjetnom inteligencijom još ne raspolažemo, softver koji se trenutačno opisuje kao AI već ima značajan utjecaj na digitalni medijski i javni prostor. Može značajno uštedjeti radno vrijeme istraživačima, novinarima, dizajnerima, programerima i drugim kreativnim profesionalcima, a poneka radna mjesta su zbog brzine razvoja velikih jezičnih modela već na rubu nestajanja; takav je, recimo, slučaj s prevoditeljima. U većini zadataka još uvijek ne mogu konkurirati ljudskim profesionalcima, između ostalog zbog učestalih činjeničnih pogrešaka (tzv. halucinacija), ali već sada su na dovoljno visokoj razini da korisnici interneta svakodnevno nailaze na sadržaj proizveden uz pomoć generativnih AI alata.

Ovaj pojam - generativni - također zahtjeva da ga se definira. U knjizi *Generative Deep Learning*, David Foster [kategorizira umjetnu inteligenciju \(AI\)](#) u dvije široke skupine: diskriminativno i generativno modeliranje. Diskriminativni modeli usredotočeni su na predviđanje odnosa između ulaza i izlaza, poput prepoznavanja objekata na slikama. Njihova svrha je automatizacija donošenja određenih

odluka temeljem podataka s kojima model raspolaže. Nasuprot tome, generativni modeli osmišljeni su za stvaranje novih podataka koji nalikuju ulaznim podacima, primjerice generiranje slika koje su teoretski potpuno nove, ali zapravo su derivati sadržaja na kojem je model treniran.

Nedavni napredak u generativnim modelima doveo je do onoga što se danas često naziva “sintetičkim medijima”. Termin sintetički mediji odnosi se na sadržaj koji je stvoren ili izmijenjen korištenjem AI alata.

Važno je uzeti u obzir i negativne aspekte. Jedan od njih je tzv. *AI slop*, internetski sadržaj vrlo niske kvalitete kojega je stvorila umjetna inteligencija, a koji je preplavio društvene mreže. Ključni problem nastaje kada se sami AI modeli treniraju na ovom sadržaju, što dodatno utječe na kvalitetu njihovih rezultata.

Povećana sofisticiranost, dostupnost i pristupačnost generativnih AI alata uvodi nove, manje očite oblike zloupotrebe. Ti oblici nisu izrazito zlonamjerni niti izravno krše uvjete korištenja alata, ali i dalje imaju zabrinjavajuće etičke implikacije. To uključuje nove načine komunikacije u političkoj promidžbi, samopromociji i zagovaranju, koji zamagljuje granicu između autentičnosti i obmane. U rukama malicioznih aktera isti ti alati postaju sredstvo za proizvodnju manipulacija, dezinformacija te štetnih pa i protuzakonitih sadržaja.

Taksonomija taktika zloupotrebe generativne umjetne inteligencije (AI), kako je prikazana u radu [“Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data”](#) autora-ica [Nahema Marchal, Rachel Xu., Rasmi Elasmr, Iason Gabriel, Beth Goldberg and William Isaac](#), razlikuje iskorištavanje sposobnosti sustava umjetne inteligencije i kompromitiranje samih sustava.

Iskorištavanje sposobnosti generativne AI

Obuhvaća tri podkategorije zloupotrebe generativne AI, od kojih je svaka povezana s određenim taktikama.

1 / Realističan prikaz ljudskog lika

Ova kategorija uključuje slučajeve u kojima sadržaji koje stvara generativna AI nastoje oponašati ljude. **“Lažno predstavljanje”** odnosi se na sadržaje koji prikazuju stvarnu osobu i pokušavaju djelovati u njezino ime u stvarnom vremenu (npr. korištenjem tuđeg lika tijekom video poziva). Sadržaji koji prikazuju stvarnu osobu u statičnom obliku (npr. lažna fotografija slavne osobe) spadaju u kategoriju **“Prisvojenog lika”**. Kada generativni AI stvara potpuno sintetičke osobe i koristi ih za djelovanje u svijetu, to se naziva **“Sock Puppeting”** (lažno predstavljanje s namjerom manipulacije). Autori smatraju da se materijali seksualno eksplicitnog sadržaja bez pristanka odraslih (NCII) i materijali seksualnog zlostavljanja djece (CSAM) trebaju tretirati kao zasebna kategorija zbog njihove “izrazito štetne prirode”, iako se u njihovom prikupljanju i distribuciji može koristiti bilo koja od tri spomenute taktike.

U slučajevima realitičnog prikaza ljudskog lika, možemo vidjeti različite tehnike kao što su:

- **Manipulacija značajkama lica:** Ova tehnika omogućuje promjenu specifičnih karakteristika lica (npr. boje očiju, oblika nosa ili strukture čeljusti) koristeći AI modele, dok ostatak lica ostaje nepromijenjen.
- **Zamjena lica (*Face swapping*):** Tehnika u kojoj se lice jedne osobe zamjenjuje licem druge osobe u fotografijama ili videozapisima.

- **Rekonstrukcija lica i sinkronizacija usana:** Ova tehnika uključuje korištenje pokreta lica i izraza jedne osobe kako bi se animiralo lice druge osobe, uključujući sinkronizaciju usana s audio zapisima ili govorom, stvarajući iluziju da osoba izgovara riječi koje zapravo nikada nije rekla.

2 / Realističan prikaz ne-ljudskih objekata

Ova podkategorija obuhvaća taktike koje koriste mogućnosti generiranja zvuka i slike za stvaranje realističnih prikaza pjesama, mjesta, knjiga i sl. **“Kršenje intelektualnog vlasništva”** odnosi se na sadržaje koje stvara generativna AI, a koji dupliciraju ljudske kreacije bez dozvole. **“Krivotvorina”** označava izlaze koji imitiraju originalna djela pretvarajući se da su autentični, dok **“falsifikacija”** nastupa kada generativni AI prikazuje lažne događaje, mjesta i predmete kao stvarne.

3 / Upotreba generiranog sadržaja

Zlonamjerni akteri mogu koristiti sposobnosti generativne AI za širenje laži u velikim razmjerima. Na primjer, mogu koristiti LLM-ove za stvaranje i širenje dezinformacija tijekom izbornih procesa kako bi utjecali na rezultate glasanja. Ova taktika se naziva **“Skaliranje i amplifikacija”**. Kada se lažnim sadržajem targetira određene publike pomoću generativne AI, ta se taktika naziva **“Ciljanje i personalizacija”**. Na primjer, jedno istraživanje otkrilo je mrežu od najmanje 64 X/Twitter botovskih računa koji koriste sadržaj generiran od strane ChatGPT-a za širenje dezinformacija ili manipulaciju javnim mnijenjem.

	Tactic	Definition	Example
Realistic depictions of human likeness	Impersonation	Assume the identity of a real person and take actions on their behalf	AI robocalls impersonate President Biden in an apparent attempt to suppress votes in New Hampshire
	Appropriated Likeness	Use or alter a person's likeness or other identifying features	Photos of detained protesting Indian wrestlers altered to show them smiling
	Sockpuppeting	Create synthetic online personas or accounts	Army of fake social media accounts defend UAE presidency of climate summit
	Non-consensual intimate imagery (NCII)	Create sexual explicit material using an adult person's likeness	Celebrities injected in sexually explicit "Dream GF" imagery
	Child sexual abuse material (CSAM)	Create child sexual explicit material	Deepfake CSAI on sale on Shopee
Realistic depictions of non-humans	Falsification	Fabricate or falsely represent evidence, incl. reports, IDs, documents	AI-generated images are being shared in relation to the Israel-Hamas conflict
	Intellectual property (IP) infringement	Use a person's IP without their permission	He wrote a book on a rare subject. Then a ChatGPT replica appeared on Amazon
	Counterfeit	Reproduce or imitate an original work, brand or style and pass as real	Fraudulent copycats of Bard and ChatGPT appear online
Use of generated content	Scaling & Amplification	Automate, amplify, or scale workflows	Researchers use GPT-3 to mass email state legislators, signaling rising verisimilitude of AI-generated emails
	Targeting & Personalisation	Refine outputs to target individuals with tailored attacks	WormGPT can be used to craft effective phishing emails

Kompromitiranje generativnih AI sustava

Uključuje taktike usmjerene na iskorištavanje ranjivosti u samim AI alatima, a ne njihovih sposobnosti, pri čemu se nanosi šteta sustavu. Napadi na integritet modela uključuju taktike koje manipuliraju modelom, njegovom strukturom, postavkama ili unosima.

- Napadi na integritet podataka uključuju taktike koje mijenjaju podatke korištene za treniranje modela ili ugrožavaju njegovu sigurnost i privatnost.
- “Steganografija” skriva kodirane poruke unutar generiranog sadržaja radi prikrivene komunikacije.
- “Trovanje podataka” korumpira skupove podataka za treniranje kako bi se u model uvele ranjivosti ili izazvale netočne predikcije.

- Napadi na “kompromitiranje privatnosti” otkrivaju osjetljive osobne informacije poput medicinskih zapisa korištenih za treniranje modela.
- Napadi “eksfiltracije podataka” nadilaze povredu privatnosti i odnose se na nezakonito dobivanje podataka korištenih za treniranje modela.
- Napadi “ekstrakcije modela” djeluju slično kao ekfiltracija podataka, ali ciljaju na sam model, pokušavajući pribaviti njegovu arhitekturu, parametre ili hiper-parametre.

	Tactic	Definition	Example
Model integrity	Prompt injection	Manipulate model prompts to enable unintended or unauthorised outputs	ChatGPT workaround returns lists of problematic sites if asked for avoidance purposes
	Adversarial input	Add small perturbations to model input to generate incorrect or harmful outputs	Researchers find perturbing images and sounds successfully poisons open source LLMs
	Jailbreaking	Bypass restrictions on model's safeguards	Researchers train LLM to jailbreak other LLMs
	Model diversion	Repurpose pre-trained model to deviate from its intended purpose	We Tested Out The Uncensored Chatbot FreedomGPT
	Model extraction	Obtain model hyperparameters, architecture, or parameters	ChatGPT Spills Secrets in Novel PoC Attack
	Steganography	Hide message within model output to avoid detection	Secret Messages Can Hide in AI-Generated Media
	Poisoning	Manipulate a model's training data to alter behaviour	Researchers plant misinformation as memories in BlenderBot 2.0
Data integrity	Privacy compromise	Compromise the privacy of training data	Samsung bans use of ChatGPT on corporate devices following leak
	Data exfiltration	Compromise the security of training data	Researchers find ways to extract terabytes of training data from ChatGPT

Na temelju te taksonomije, [istraživači su analizirali “medijske izvještaje o zloupotrebi generativne AI između siječnja 2023. i ožujka 2024.”](#). Konkretno, autori su proučavali popularnost određenih taktika i ciljeve povezane s njihovom primjenom.

Glavni nalaz studije pokazao je da je 9 od 10 dokumentiranih slučajeva uključivalo iskorištavanje sposobnosti generativne AI za zloupotrebu. Najčešće korištene taktike manipulirale su ljudskim likom, prvenstveno putem lažnog predstavljanja i *sockpuppetinga*. Većinom su primjenjivane s jasnim ciljem utjecanja na javno mišljenje, omogućavanja prevara ili nezakonitih aktivnosti te stvaranja profita.

Najčešći cilj iskorištavanja generativnih AI sposobnosti bio je utjecanje na javno mišljenje pomoću *deepfakeova*, digitalnih persona i drugih falsifikata (27% svih prijavljenih slučajeva). Korištenje generativne AI za skaliranje, amplifikaciju i monetizaciju sadržaja također je bilo vrlo često (21% svih slučajeva), dok su prevare činile 18% svih prijavljenih slučajeva. Zanimljivo, korištenje generativne AI za maksimalizaciju dosega bilo je relativno rijetko (3,6% prijavljenih slučajeva).

U većini slučajeva zloupotrebe AI sustava primarno se radi o iskorištavanju lako dostupnih mogućnosti generativne AI, koje zahtijevaju minimalnu tehničku stručnost. Zbog te jednostavnosti korištenja, AI ima važnu ulogu i u proizvodnji i širenju dezinformacija. Sustavi umjetne inteligencije omogućuju zlonamjernim akterima da jednostavno stvaraju velike količine relativno uvjerljivog lažnog sadržaja te pomažu njegovom brzom širenju.

AI dezinformacije su činjenično netočne, neutemeljene i manipulativne tvrdnje u tekstualnom ili audiovizualnom formatu koje su generirane korištenjem alata umjetne inteligencije.

Kada govorimo o AI dezinformacijama, razlikujemo dva načina nastanka:

- **Digitalno izmijenjene dezinformacije** se odnose na bilo koji sadržaj koji je značajno izmijenjen kako bi se manipuliralo ili

promijenilo poruku koju je izvorno prenosio. To ne uključuje izmjene radi jasnoće ili kvalitete.

- **Dezinformacije generirane umjetnom inteligencijom** se odnose na bilo koji oblik “novog” sadržaja koji je stvorio sustav umjetne inteligencije.

AI generirani
sadržaj u SEE regiji

Portal Fake News Tragač objavio je u veljači 2025. analizu [“GOING VAIRAL - Viralnost sadržaja kreiranih pomoću AI”](#). Analizom su obuhvatili podatke o detektiranim AI dezinformacijama u periodu od siječnja do listopada 2024. godine, a koje su prikupile organizacije Fake News Tragač, Istinomer, Raskrikavanje i AFP činjenice iz Srbije, Raskrinkavanje.ba iz Bosne i Hercegovine, Raskrinkavanje.me iz Crne Gore te Faktograf iz Hrvatske. Autori analize Stefan Janjić i Marina Grnja Klaić navode kako je od 88 analiziranih AI generiranih sadržaja, najviše razotkrivenih AI dezinformacija i manipulacija zabilježeno na portalima Faktograf.hr i Raskrinkavanje.me. Vremenska distribucija objava ne pokazuje jasnu pravilnost – iako su neke bile vezane za aktualne događaje (proteste, ratove), uglavnom su predstavljale dekontekstualizirane manipulacije koje se lako mogu reciklirati u različitim vremenskim periodima. Većina objava odnosila se na javne osobe, ali bez direktne povezanosti s trenutnim zbivanjima. Najviše detektiranog i obrađenog AI generiranog sadržaja pronađeno je na Facebooku, što je i logično s obzirom na to da je većina navedenih organizacija dio Metina *Third Party Fact-Checking* programa. Od vrsta AI manipulacija, najmanje je onih tekstualnih, a dominiraju AI fotografije te AI video sadržaji. Autori analize posebno ističu da, iako su vizualne AI manipulacije u njihovom uzorku zastupljenije od tekstualnih, to ne znači da ih u stvarnosti ima manje, već da su teže uočljive i dokazive. Naime, alati kojima se mogu detektirati lažne fotografije i video sadržaji, iako su daleko od nepogrešivih, trenutno su ponešto pouzdaniji od onih koji detektiraju AI generirani tekst. Autori zaključuju da analiza pokazuje da se AI dezinformacije i manipulacije šire kroz tri ključna obrasca:

- **Brzina** – nastaju i šire se u minutama, dok provjere točnosti i moderacija sadržaja kasne, što stvara „ranjive periode” za masovni doseg.

- **Skalabilnost** – jednom uspješan format lako se umnožava u stotinama varijacija.
- **Emocionalna manipulacija** – izazivaju snažne reakcije publike i često prolaze ispod radara algoritama jer ne krše direktno pravila platformi.

Za potrebe ove Faktografove analize prikupljene su informacije o AI generiranom sadržaju članica SEE Check mreže i njihovih portala [Raskrinkavanje.me](https://raskrinkavanje.me), [Raskrinkavanje.ba](https://raskrinkavanje.ba), [FakeNews Tragač](https://fakenews.tragac.hr), [RasKRiKavanje](https://raskrinkavanje.hr) i [Faktograf.hr](https://faktograf.hr), a koje se bave monitoriranjem i razotkrivanjem dezinformacija, manipulacija i lažnih vijesti. Zanimalo nas je koliko se uočenog sadržaja odnosi na lokalne aktere i kontekste, a koliko na međunarodne, kao i na koje se teme taj sadržaj odnosi te pojavljuje li se u zemljama regije politički AI generirani sadržaj. Prikupljene informacije ne predstavljaju sveobuhvatan prikaz AI generiranog sadržaja u SEE regiji već proizlaze iz svakodnevnog rada portala i njihova monitoriranja sadržaja na društvenim mrežama.

Hrvatska

Tijekom svog rada, a posebice u posljednje dvije godine, Faktograf je detektirao više od 60 AI generiranih sadržaja u obliku dezinformacija, manipulacija i lažnih vijesti. Kao što je i navedeno u ranije spomenutoj analizi kolega iz Fake News Tragača, radi se o sadržajima koji se najčešće pojavljuju u obliku videa i slika, dok je tekstualnih primjera zanemarivo malo, a isključivo audio sadržaji nisu zabilježeni.

U Faktografovom uzorku najprisutnije su dezinformacije vezane za Hrvatsku (40%), a potom slijede i druge regije svijeta kao što su SAD, Bliski istok i Europa. U slučaju lokalnih AI generiranih dezinformacija,

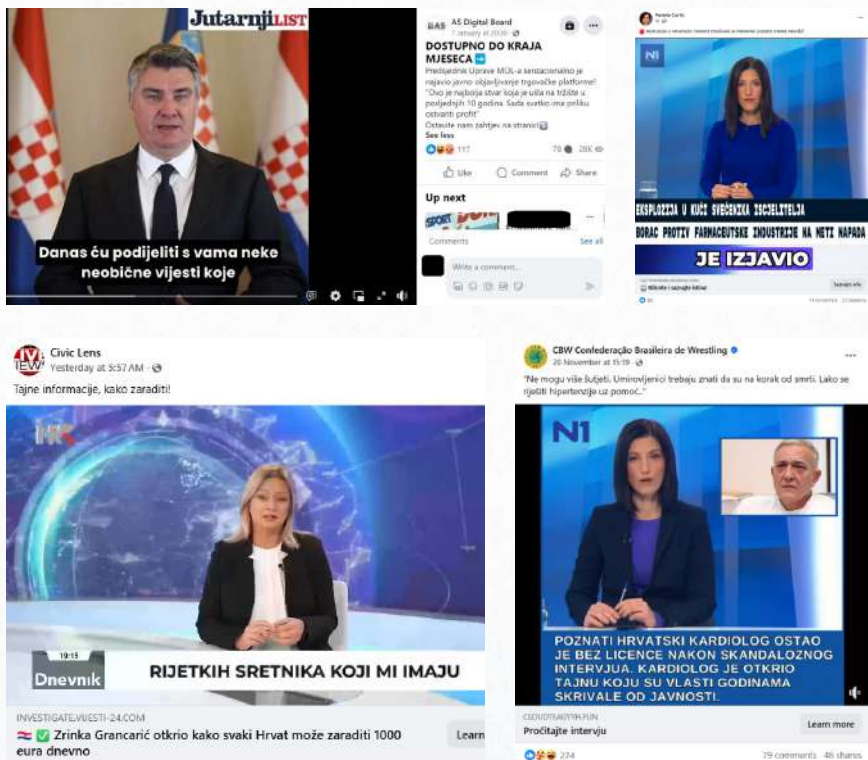
one se gotovo isključivo odnose na financijske i zdravstvene prevare. Osim toga, među hrvatskim se korisnicima često pojavljuju AI generirane manipulacije povezane s inozemnom politikom, ratovima te elementarnim nepogodama. To ukazuje da se manipulativni i lažni sadržaji pojavljuju u vrijeme kriza i u vezi s društveno osjetljivim temama te služe promoviranju narativa o tim događanjima (primjerice izazvanju sućuti prema žrtvama rata ili diskreditiranju političara).



Izvor: Faktograf.hr

AI generirani sadržaji koji se odnose na prevare najčešće zloupotrebljavaju fotografije i video snimke [novinara](#), [političara](#), [državnih dužnosnika](#) i [znanstvenika](#), posebice liječnika. Prevaranti kreiraju *deepfake* video sadržaje koji simuliraju ili koriste televizijske emisije odnosno novinarske intervjuje, a gosti, političari ili znanstvenici u njima govore o mogućnostima za brzu zaradu ili magičnim lijekovima za raširene zdravstvene poteškoće. Cilj takvih sadržaja

je nanijeti štetu građanima koristeći javne osobe kojima oni vjeruju, što potencijalno može dovesti do slabljenja povjerenja u ključne društvene i demokratske aktere.



Izvor: Faktograf.hr

Ako promatramo specifično izborna razdoblja, Faktograf u svojem izvještaju o dezinformacijama tijekom lokalnih izbora 2025. - “Činjenice iza kampanje” - navodi kako su na početku superizborne 2024./2025. godine postojale bojazni o mogućem utjecaju dezinformacija generiranih pomoću umjetne inteligencije na izborni proces, ali to se nije ostvarilo.

AI generirani audio i video politički sadržaj u pravilu čine dvije velike skupine. Prva su prevare koje koriste AI generirane snimke političara, o čemu je već bilo riječi, a drugo su satira, ismijavanje, politički komentar i politička propaganda, pri čemu se nerijetko pribjegava difamacijama, stereotipima pa i širenju mržnje, posebno prema migrantima, stranim radnicima i LGBTIQ+ zajednici. Vrlo često AI kreirani video sadržaj nije jasno označen, no iz tona i narativa koji se u tim videima koriste, jasno je da se radi o političkoj propagandi, a ne pokušaju da se birače uvjeri da im se prikazuju stvarni događaji.



nardnjeporci
100,000 views

Uspostili ste stvarni prijevrat? AI generirane "Narodnjeporci" i "Alena Filipović" - Dva puta više puta i 100 puta više puta, jer vam u ovom videu prikazujemo 11 glavnih događaja iz života Alene Filipović, glavni događaji u životu i spomeni na njezin život (Alena Filipović), spomeni na njezin život (Alena Filipović) i njezin život (Alena Filipović).



Izvor: Faktograf.hr

Bosna i Hercegovina

[Raskrinkavanje.ba](https://www.raskrinkavanje.ba) navodi da su se AI generirani sadržaji koje su analizirali odnosili kako na lokalne aktere i lokalni kontekst, tako i na inozemne aktere i događaje. Kao i u slučaju drugih organizacija iz SEE regije, susretali su se sa sadržajem koji se odnosio na međunarodno poznate javne osobe, ratne sukobe, prirodne pojave i katastrofe te na druge globalne događaje koji su u fokusu javnosti.

“Iako smo se susretali sa lokalnim primjerima AI dezinformacija koje imaju političku ili društvenu konotaciju, većina naših analiza u tom kontekstu vezana je za promotivne kampanje za zdravstvene prevare. To su uglavnom *deepfake* videosnimci u kojima se zloputrebljavaju popularne ličnosti iz zemlje i regiona, putem kojih se prodaju razni preparati koji obećavaju zdravstvene benefite ili liječenje”, navodi Amar Karađuz, novinar portala [Raskrinkavanje.ba](https://www.raskrinkavanje.ba).



Izvor: [Raskrinkavanje.ba](https://www.raskrinkavanje.ba)

Uz AI generirane medicinske prevare, [Raskrinkavanje.ba](#) se bavilo i [YouTube kanalom Crna Hronika TV](#) koji je objavljivao AI generirane sadržaje. Između ostalog, velik broj sadržaja objavili su o slučaju djevojčice Danke Ilić koja je ubijena početkom 2024. godine u Srbiji, pri čemu su tvrdili da je biološki otac djevojčice iz BiH te da postoji mogućnost da je djevojčica kod njega, a plasirali su i druge lažne informacije o tom slučaju. U proljeće 2024. godine, kada je [Raskrinkavanje.ba](#) radilo analizu sadržaja, kanal je imao 9300 pratitelja s ukupno 2,4 milijuna pregleda sadržaja.



Izvor: [Raskrinkavanje.ba](#)

Još jedan primjer lokalno nastalih AI generiranih sadržaja pokazuje kako pojavu i širenje takvih sadržaja mogu inspirirati stvarni događaji. Nakon napada na policijsku stanicu u Bosanskoj Krupi u jesen 2024. godine, radi čega su uhapšeni petnaestogodišnji napadač i grupa osoba koja se dovodila u vezu s napadom, među korisnicima društvenih mreža u regiji [kružila je fotografija navodnog hapšenja pripadnika selefijskog pokreta](#) uz jasne islamofobne poruke.



Izvor: Raskrinkavanje.ba i Raskrinkavanje.me

“U BiH nije primijećeno značajnije korištenje AI generiranog sadržaja u svrhe političke propagande, ali budući da se ove godine očekuju referendum i prijevremeni izbori u Republici Srpskoj, kao i da bi se naredni Opći izbori trebali održati 2026. godine, upravo će ovaj sadržaj biti u fokusu rada Raskrinkavanja.ba”, zaključuje Amar Karađuz.

Crna Gora

Kako je navedeno u analizi FakeNews Tragača, portal Raskrinkavanje.me je, uz Faktograf, razotkrio najveći broj AI generiranog sadržaja koji cirkulira među korisnicima društvenih mreža, primarno Facebooka, u SEE regiji. Jelena Jovanović, izvršna urednica portala, navodi kako većina tog sadržaja nije kreirana u Crnoj Gori, već je riječ o sadržaju koji je uglavnom preuzet sa regionalnih dezinformativnih stranica i profila na društvenim mrežama, ali i pojedinih rubnih portala poznatih

po širenju različitih teorija zavjera. Sadržajno se često radi o političkim temama, ratovima, ali i manipulacijama vezanih za elementarne nepogode i klimatske promjene.



Izvor: Raskrinkavanje.me

I u Crnoj Gori korisnici društvenih mreža susreću se s različitim prevarama. Primjerice, primjećene su prevare kojima su se nudile [besplatne karte za gradski javni prijevoz u Podgorici](#), [tvrdilo se da](#)

korisnici mogu dobiti pogodnosti za kupovinu goriva na određenim benzinskim stanicama, ili “gotovo besplatne” kućanske aparate.



Izvor: Raskrinkavanje.me

Kada su u pitanju prevare, i u crnogorskom slučaju dolazi do zlouporabe lica političara, poput predsjednika i premijera, ali i predsjednika Skupštine, čije se fotografije, kao i kreirani *deepfake* snimci, zloupotrebljavaju u cilju online prevara za ulaganje u kriptovalute.

“Imamo i primjere ljekara, koji su bez pristanka i znanja postali protagonisti marketinških kampanja na internetu. Riječ je o *deepfake* snimku na kojem jedan poznati nikšički kardiolog navodno preporučuje određene preparate za erekciju. Osim njega, meta ovakvih reklama bili su i pojedini voditelji i novinari koji su navodno reklamirali preparate za mršavljenje, kao i inovacije u nauci za otkrivanje načina da se sačuva vid”, navodi Jelena Jovanović.



Izvor: Raskrinkavanje.me

U slučaju političke propagande i poticanja nacionalne netrpeljivosti, Raskrinkavanje.me zabilježio je [primjer zlouporabe jednog satiričnog AI snimka](#) koji je iskorišten za međuvjerska i međunacionalna “prepucavanja” na društvenim mrežama. “U minut i nešto prikazan je navodni prilog o tome kako Bošnjaci shvataju da su u stvari Srbi i da je imalo ‘smisla da se vrate svojim korijenima’. Iako je dosta korisnika društvenih mreža koji su vidjeli ovu objavu ukazalo na to da se radi sadržaju kreiranom uz pomoć vještačke inteligencije (AI), u komentarima se ipak povela žučna rasprava o islamu i hrišćanstvu”, navodi Jovanović.

Raskrinkavanje.me detektiralo je i [primjer političkog AI generiranog sadržaja](#) u kojem je jedan korisnik objavio navodnu snimku srpskih studenata, tzv. Blokadera, kako pohađaju Thompsonov koncert u Zagrebu, što odgovara narativima diskreditacije studenata kao hrvatskih agenata, slično kao i u primjeru koji je detektiralo RasKRIKavanje, a navodimo ga niže u tekstu.



Dušan Ampel Dimitrijević · Follow

July 7 at 6:28 AM · 🌐

...

Blokaderi na Thompson-ovom koncertu, naravno. ✅



👍👎👉 252

289 comments 80 shares

Izvor: Raskrinkavanje.me

Srbija

U RasKRIKavanju navode da su tijekom svojeg rada uglavnom bilježili AI sadržaj lokalnog karaktera, odnosno ticao se aktualnosti u Srbiji. Primjećuju da se AI najčešće koristi za manipulaciju fotografijama. No, RasKRIKavanje izdvaja primjer [tekstualne AI manipulacije](#), koju su uočili na stranici jednog tabloida, o baki koja je navodno nasamarila

lopove. Ispostavilo se da je ta priča stara dvije godine i da se događaj desio u Americi. “Uočili smo da je ‘domaća’ verzija teksta - koja je bila izmenjena u odnosu sa stvarni događaj - napisana pomoću veštačke inteligencije jer su korišćene čudne jezičke konstrukcije koje nisu u duhu srpskog jezika”, navodi Milica Ljubičić, novinarka RasKRIKavanja.

Najčešće teme uočenih AI manipulacija, u njihovom slučaju, ticale su se uglavnom politike. Krajem 2024. godine zabilježili su [AI generiranu sliku Donalda Trumpa koji drži navodno poderanu zastavu Kosova](#). Tako nešto se nije dogodilo, a vidljive nepravilnosti na fotografiji su jasno ukazivale da je za tu fotografiju korišten neki AI alat.

“Takođe smo u aprilu ove godine pisali o manipulaciji koju je plasirao poslanik vladajuće stranke i to tako što je objavio [AI generisanu fotografiju sa namerom da diskredituje studente u blokadi](#)”, navodi Ljubičić te dodaje kako se radi o očitom obliku političke propagande. Radilo se o AI generiranoj fotografiji koja prikazuje studente za stolom kako jedu rukama, dok iznad njih visi velika hrvatska zastava. “Ova objava je za samo nekoliko sati došla do preko 11.000 pregleda, 120 deljenja i više od 200 lajkova”, navodi se u tekstu na RasKRIKavanju.



Izvor: RasKRIKavanje

Iz RasKRIKavanja ističu kako su se ranije na Facebooku pojavljivali i AI generirani videi sa svrhom financijske prevare građana, a u kojima navodno Novak Đoković poziva građane da ulažu u kriptovalute i tako zarade.

Teodora Koledin, novinarka Fake News Tragača, navodi kako su se u radu susretali i s lokalnim i s globalnim AI generiranim obmanjujućim i lažnim sadržajima. Kada se radi o lokalnim lažnim sadržajima, uglavnom su u pitanju različiti *deepfake* video zapisi poznatih osoba (primjerice [Branimir Nestorović](#), [Ana Đurić Konstrakta](#)) kreiranih u svrhe prodaje neprovjerenih medicinskih preparata. Koledin navodi da su ovi *deepfake* videi na Facebooku oglašavani tako da ciljaju osobe starije od 65 godina. Dodatno su se, kao i kolege iz RasKRIKavanja, susretali i s financijskim prevarama.



Izvor: FakeNews Tragač

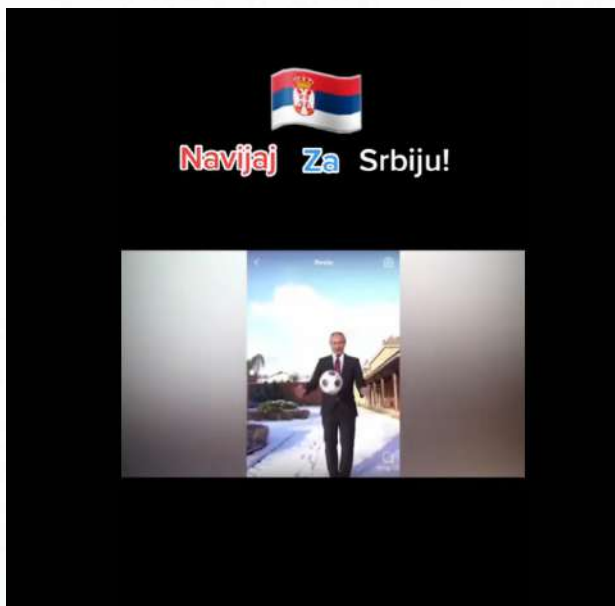
Također, kako navode, na pojedinim nacionalnim televizijama djeljeni su [AI generirani sadržaji s krivotvorenim izjavama opozicijskih srpskih političara](#), pri čemu često ne bi postojala napomena o korištenju AI-a za nastanak tih sadržaja. “No, čak i kada uključuje napomenu, vrlo je upitno koliko je ona dovoljna da publici ovih medija ukaže na neautentičnost prikazanih audio-vizuelnih snimaka”, ističe Koledin.

Lažni sadržaji globalnog karaktera također su se našli u programima nacionalnih medija pa tako Koledin navodi [primjer Hepi televizije](#): “U jednoj je emisiji podeljen *deepfake* aktivistkinje Grete Thunberg u kojoj ona – navodno – podstiče na korišćenje ‘veganskih granata’ i ‘biorazgradivih raketa’. Čak i u trenutku kada gost programa preispituje autentičnost snimka, voditelj ga uverava da je on verodostojan”.



Izvor: FakeNews Tragač

Jedan su AI generirani video kao stvaran prikazali i [Portal NSuživo te Portal Srpskog telegrafa Republika](#), objavivši kako će ruski predsjednik Vladimir Putin podržati nogometnu reprezentaciju Srbije na Svjetskom prvenstvu u nogometu.



Izvor: TikTok

Narativi o Al-ju u SEE regiji

Koristeći monitoring-analitički sustav *Gerluata Juno* analizirali smo AI narative koji cirkuliraju društvenim mrežama u SEE regiji. Ograničenje ove analize je u tome što je nemoguće obuhvatiti sav sadržaj dostupan na društvenim mrežama. U monitoring sustav unijeli smo dio ključnih stranica i kanala na društvenim mrežama koje organizacije aktivne u SEE Check mreži redovito prate. Dodatno, treba imati na umu da sustav može pratiti samo Facebook stranice, ali ne i Facebook profile, koji često dominiraju među akterima koji šire manipulativan i dezinformirajući sadržaj, te da prati samo javno dostupni sadržaj.

Analizom smo obuhvatili više od 1000 jedinica sadržaja objavljenih na više od 50 različitih stranica i kanala na društvenim mrežama Telegram, TikTok, Instagram, Youtube i Facebook, u periodu od ožujka do kolovoza 2025. godine. S obzirom na to da u uzorku dominiraju kanali i stranice na Telegramu i TikToku, ne čudi i da je najveći broj sadržaja pronađen na praćenim Telegram kanalima. Ipak, iako je na TikToku detektiran manji broj sadržaja u odnosu na Telegram, on postiže značajnu vidljivost. Tako podaci iz sustava pokazuju da dok je sa 912 jedinica sadržaja na Telegram kanalima interakciju imalo 65 640 korisnika te je procijenjeno da je sadržaj pogledan 2.393 milijuna puta, sa 64 jedinice sadržaja na TikToku interakciju je imalo 48 610 korisnika te je sadržaj pregledan 1.48 milijuna puta. To znači da je 14 puta manje sadržaja na TikToku generiralno samo 1,3 puta manje interakcija i 1,6 puta manje pregleda. Razlika proizlazi i iz različitog načina funkcioniranja te dvije mreže, ali pokazuje da sadržaj objavljen na TikToku ima veći potencijalni doseg do korisnika od sadržaja u Telegram kanalima.

Analiza je pokazala da se glavni narativi vezani za AI odnose na tehnološke, geopolitičke, ekonomske i društvene učinke AI-a, a objave se kreću od vijesti i analiza do teorija zavjere i spekulativnih komentara. Rasprave o AI-u se isprepliću sa širim temama poput

globalnog upravljanja, gubitka privatnosti, vojne modernizacije, rada i zapošljavanja te društvenih transformacija. Česte su zabrinutosti zbog dezinformacija, manipulacija i egzistencijalnih rizika koje AI donosi, posebno *deepfakeova* i automatiziranog lažnog predstavljanja, kao i regulatornih, etičkih i kulturnih pitanja.

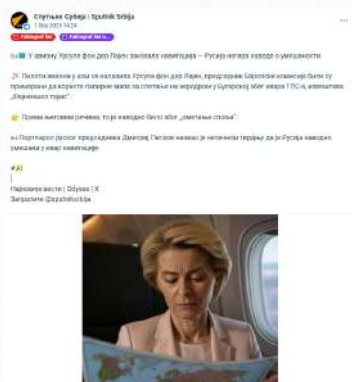
Ključne teze koje se pojavljuju vezane za AI mogu se sažeti u nekoliko točaka:

- AI tehnologije se doživljavaju istodobno kao pokretač napretka i izvor dubokih društvenih, ekonomskih i egzistencijalnih rizika.
- Prisutna je svijest o rizicima manipulacija pokretanih i generiranih umjetnom inteligencijom kao što su deepfakeovi, lažne vijesti, lažna predstavljanja i slični, što može dovesti do povećanja nepovjerenja u medije i društvo u cjelini.
- Geopolitičke napetosti sve više uključuju i aspekt AI-a, osobito u kontekstu odnosa između globalnih sila te sigurnosti.
- Postoji zabrinutost oko masovne automatizacije rada, gubitka radnih mjesta i transformacije tradicionalnog zapošljavanja, uz konkretna pitanja o tome koja će zanimanja opstati.
- Prisutna je skepsa i prema zapadnim i prema kineskim AI modelima, uz rasprave o lokalizaciji, suverenitetu i regulatornom nadzoru.

Glavne teme koje se problematiziraju na promatranim stranicama i kanalima mogu se prikazati u nekoliko tematskih grupa.

AI kao geopolitički i vojni čimbenik

- AI se smatra strateškim sredstvom koje oblikuje globalne odnose moći, kako one trgovinske, tako i vojne te utječe na ishod sukoba na razini država, osobito među globalnim silama.
- AI se koristi u ratovanju i sigurnosti - napredak u autonomnim dronovima, kibernetičkom ratovanju i obavještajnim kapacitetima.
- AI ima svoje mjesto i u diplomaciji i upravljanju kao alat u pregovorima, mirovnim inicijativama i stvaranju novih međunarodnih saveza ili pravnih okvira.



ND Novo doba starih ratnika
21 Aug 2025 11:58

Faktograf list **HR** **Faktograf list u...**

Izvršni direktor Palantira Alex Karp: Naša **umjetna inteligencija** igra sve 'ključniju ulogu' u ratu

"Nastavit ćemo igrati ključnu ulogu u toj areni i usput, putem našeg softvera omogućujemo drugim softverskim i hardverskim tvrtkama da rade na našoj platformi, tako da ovo nije samo pobjeda za Palantir, iako jest i jako sam ponosan na to, ovo je, iskreno rečeno, pobjeda za tehnološki ekosustav koji pokreće ovu naciju."
<https://t.me/Novodobastarihratnika>

@ [http://t.me/Novodobastarihratnika](https://t.me/Novodobastarihratnika)

@ t.me

16 interactions 735 views

Izvor: Gerluata Juno

AI generirane dezinformacije, deepfakeovi i manipulacije

- Upozorava se na opasnost od širenja *deepfakeova* i ilegalnog sadržaja
- Uočavaju se političke manipulacije tijekom izbora te politička propaganda kreirana AI alatima, pri čemu se na AI gleda kao potencijalni alat društvenog inženjeringa i manipulacije javnog mijenja putem bot-farmi, AI generiranih objava i lažnih računa za stvaranje *online* atmosfere koja ne odgovara realnosti.
- Ističe se važnost prepoznavanja AI generiranog sadržaja te potreba za AI pismenošću.



Izvor: Gerluata Juno

Socioekonomski utjecaji AI-ja i automatizacije poslova

- Prisutan je strah od masovnog gubljenja poslova te se raspravlja o kreiranju zanimanja otpornih na utjecaj umjetne inteligencije uz rasprave o suradnji ili natjecanju čovjeka i AI-ja u radnom okruženju.
- Prisutan je i strah od stvaranja digitalnih panoptikuma i koncepta društvenih kredita što uključuje strahove od kapitalizma nadzora, digitalnih identiteta i algoritamske kontrole.
- Pojavljuju se rasprave o integraciji AI-ja u svakodnevni život i industriju, zdravstvo, financije, trgovinu, poljoprivredu.

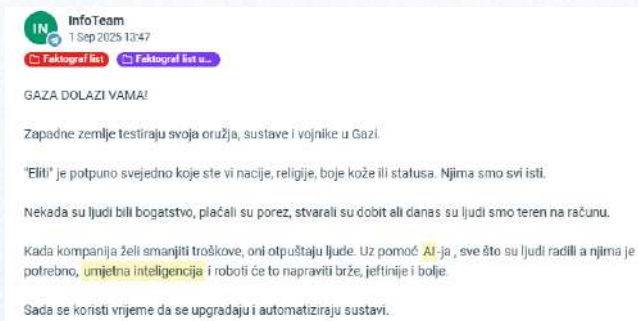




Izvor: Gerluata Juno

AI suverenitet, lokalizacija i kulturni narativi

- AI se u dijelu stranica i kanala predstavlja kao instrument zapadne hegemonije te se promovira percepcija zapadnih AI alata kao oblika nadzora i kulturnog inženjeringa.
- Zagovara se razvoj domaćih AI rješenja radi očuvanja suvereniteta i identiteta.
- Razmatra se kulturni i jezični okvir AI-ja kroz utjecaj lokalnih jezika, vrijednosti i tradicija na AI sustave.



Izvor: Gerluata Juno

Etika, filozofija i egzistencijalne rasprave

- Problemizira se koncept transhumanizma i gubitka ljudskosti; često i u kontekstu teorija zavjera.
- Postoji prisutnost tehnopesimizma i distopijskih scenarija poput digitalnog ropstva.
- Poziva se na regulaciju AI-ja te uspostavljanje sigurnosnih mjera na lokalnoj i globalnoj razini.



MT in ZA JEDNICA
23 Aug 2025 10:35

...

Faktograf list HR Faktograf list u...

TheHighwire : AI i globalna kontrola mreže

Istraživački novinar Alex Newman upozorava kako se veštačka inteligencija, transhumanizam i digitalni nadzor spajaju u globalni sistem kontrole. Od digitalnih identifikacija do CBDC-a, on upozorava da svet u srušene tone u "digitalni gulag" osmišljen da prati i manipuliše svakim pokretom.

"Ovde govorimo o nečemu što ne samo da može posmatrati i nadzirati sve što radite, već i što može početi manipulirati stvarima koje radite.

Žele jedinstveni svetski politički, ekonomski, pa čak i verski sistem. I to nam otvoreno govore već dugo vremena. Veruju da nas ima previše na planeti. Veruju da nam je bolje bez individualnih prava, da nam je bolje da nama vladaju.

Apsolutno je jasno da će pokušati iskoristiti AI tehnologiju kako bi eliminirali ljudsku slobodu, eliminirali ljudsku privatnost i u osnovi nas usmerili u pravcu koji je toliko radikalan da gotovo zvučite ludo kad o tome pričate. Ali ovi ljudi to rade."

Izvor: Gerluata Juno

Kanale i stranice na kojima su detektirani narativi o AI-ju u našem uzorku mogu se podijeliti u četiri klastera.

Prvi i dominantni klaster čine stranice i kanali konspirativnog karaktera koji često promoviraju teorije zavjere te se na njima naglašavaju različiti negativistički, pesimistični i distopijski scenariji široke upotrebe AI tehnologija. Dominacija ovog klastera u našem

uzorku je logična obzirom na prirodu medija koje organizacije iz SEE Check mreže monitoriraju. U kontekstu *mainstream* kanala, kao drugog klastera, češće se pronalaze uravnotežene rasprave koje problematiziraju efekte AI-ja na društvo i ekonomiju. Proruski kanali, kao zasebni treći klaster, češće naglašavaju vojnu upotrebu AI-ja i globalno nadmetanje u razvoju tehnologije.

Značajnu prisutnost imaju i informativno-zabavni kanali kao četvrti klaster, najčešće TikTok, Instagram i YouTube profili koji populariziraju AI kroz humor, satiru, društveni komentar i pop-kulturu, s refleksijama na lokalne kontekste, pri čemu treba imati na umu da su pojedini kanali, iako naizgled satirični, u službi političke propagande, što je bilo posebno vidljivo u tijekom hrvatskih lokalnih izbora.



Izvor: Gerluata Juno

Zaključak

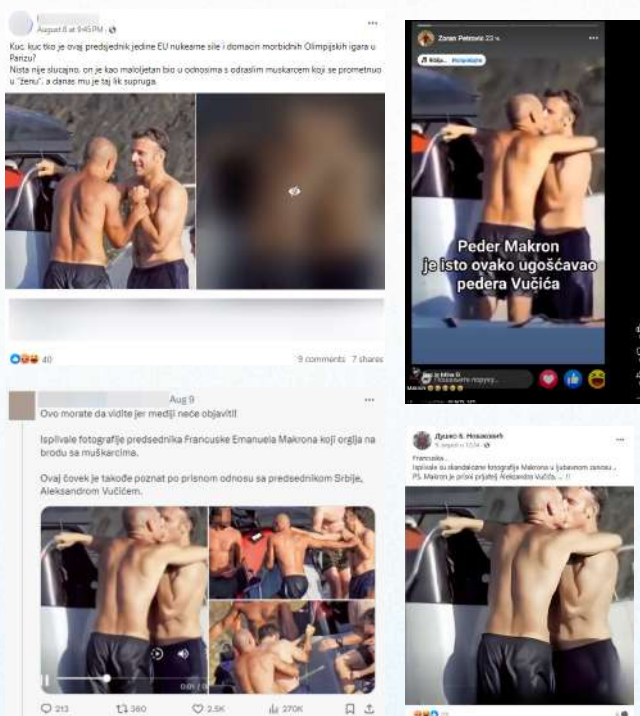
Analize koje su proveli FakeNews Tragač i Faktograf ukazuju na određene trendove vezano za manipulativni AI generirani sadržaj na području SEE regije. Najzastupljenije su vizualne manipulacije – lažne fotografije i *deepfake* videi. U svim državama obuhvaćenima analizom bilježi se zloupotreba AI sadržaja za financijske i medicinske prevare. Zajednički obrazac jest zloupotreba prepoznatljivih javnih osoba - političara, znanstvenika i liječnika - kako bi se stvorio dojam vjerodostojnosti.

Nažalost, u Hrvatskoj institucije malo toga poduzimaju kako bi zaštitile žrtve prevare, uključujući i osobe čije se ime i lice koristi za prevaru. Kako navodi Faktografov novinar Ivan Nekić u svome tekstu [Što sve hrvatske institucije \(ne\)čine po pitanju prevara generiranih umjetnom inteligencijom](#), legislativa ne prati tehnološki razvoj te unutar MUP-a ne postoji posebna statistika koja bi se mogla iskoristiti za provjeru koliko je ovakvih slučajeva uopće zabilježeno. Tijela koja prema europskoj i nacionalnoj legislativi, poput DSA-a, imaju ovlasti prijavljati nezakonit sadržaj platformama se tek upoznaju s vlastitim ovlastima, ali i u tim slučajevima radi se samo o uklanjanju sadržaja, ne i otkrivanju i kažnjavanju počinitelja. S obzirom na to da druge države bilježe sve veći broj AI generiranih kaznenih djela, jasno je da kaskanje hrvatskih institucija za tehnološkim razvojem može ugroziti sigurnost građana.

Nimalo bolja nije situacija niti u ostatku SEE regije koja, s obzirom na to da njezine države nisu članice Europske unije, nemaju niti pokrivenost postojećim europskim zakonodavstvom. O tome je opsežno za SEE Check pisala Maida Salkanović u članku [Why the Western Balkans Need the Digital Services Act](#), ističući kako je nacionalno zakonodavstvo u regiji vrlo slabo. Prema [istraživanju](#) koje je provela organizacija [Partneri za demokratske promene Srbija](#), naponi za usklađivanje nacionalnog zakonodavstva s načelima DSA-a

bili su fragmentirani i nedovoljni. Umjesto donošenja jedinstvenog regulatornog okvira koji bi sustavno obuhvatio odgovornost platformi, transparentnost algoritama i zaštitu korisnika, došlo je tek do izolirane izmjene postojećih zakona, poput onih koji se odnose na audiovizualne medije ili zaštitu potrošača.

Uz prevare, značajnu cirkulaciju u regiji imaju sadržaji koji se tiču globalnih događanja posebice vezano za političare, ratove i elementarne nepogode. Organizacije koje se bave razotkrivanjem AI generiranih manipulacija često nailaze na identične sadržaje među korisnicima društvenih mreža u pojedinoj zemlji što ukazuje na značajnu transnacionalnost dezinformacijskog sadržaja.



U regiji primjećujemo i prisutnost zabavnog, satiričnog i drugog AI generiranog sadržaja koji je ili označen kao AI ili je “na oko” vidljivo da se ne radi o autentičnom sadržaju - mada smo u više navrata primjetili da je i takav sadržaj među određenim korisnicima izazvao debate ili teorije zavjere. Pojedini mediji i kanali na društvenim medijima kontinuirano koriste AI generirane vizualne materijale za opremanje tekstova ili objave na društvenim mrežama, koje mogu ići od relativno beningnih, jeftinih ilustracija pa do ozbiljnije političke propagande koja služi kreiranju javnih narativa, a često i difamaciji političara ili sisanju mržnje protiv određenih društvenih skupina. Posebno zabrinjavaju slučajevi objave AI generiranih sadržaja u *mainstream* medijima, pri čemu se u nekim slučajevima radi o nekritičkom preuzimanju takvog sadržaja, ne provjeravajući prethodno njegovu autentičnost, ili svjesnom plasiranju takvog sadržaja bez da su gledatelji jasno informirani da se radi o AI generiranom sadržaju.

Kada se promatraju napisi o razvoju AI tehnologija, primjećujemo da stranice koje su i inače sklone teorijama zavjere, na isti način pristupaju i umjetnoj inteligenciji. Prisutni su strahovi od masovnog gubitka zaposlenja, gubitka ljudskosti, globalnog i opsežnog nadzora nad ukupnim životom ljudi te skepticizam u neutralnost AI modela koji se promatraju kao samo još jedan alat globalnih elita.

U konačnici, možemo zaključiti da SEE regija prati globalne rasprave i trendove po pitanju AI-ja, da je značajna količina sadržaja uvezena iz izvora izvan regije, no da se pojavljuju i originalno kreirani AI sadržaji koji služe političkoj propagandi lokalnih aktera te da u ovom trenutku najveću prijetnju građanima predstavljaju AI generirane financijske i medicinske prevare koje zloupotrebljavaju likove političara, novinara i znanstvenika te vizualni identitet medijskih programa.

Faktograf.hr

Facebook | Instagram | TikTok | Youtube | WhatsApp

F-zin: Faktografov newsletter



Pratite AI dezinformacije u Hrvatskoj: <https://faktograf.hr/ai-dezinformacije/>