

УЛОГА ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ У ГЕНЕРИСАЊУ ДЕЗИНФОРМАЦИЈА

Ладислав Лазић¹

Мара Скенџић Рељић²

Милијана Кочић³

Резиме: Појава вештачке интелигенције довела је до муњевитог стварања и пласирања дезинформација. Алати које користи су толико убедљиви да је све теже разликовати лажну вест од истините. "Поплава" лажних вести које се пласирају преко друштвених мрежа носи ризике од губитка поверења у медије, науку, институције, рушења угледа личности, као и опасност по ментално здравље. Са друге стране, заједно са критичким размишљањем, медијском писменошћу и етичким принципима људи, вештачка интелигенција је и савезник у борби против дезинформација. Употребом индукативне, дедуктивне, као и методе анализе и синтезе, намеће се закључак да је вештачка интелигенција оружје у људским рукама којим у два супротна правца можемо да обликујемо медијски простор.

Кључне речи: *вештачка интелигенција, лажне вести, дезинформације, медијска писменост, друштвене мреже*

¹ Ладислав Лазић, руководилац сектора куповине, продаје и размене програмског садржаја, Јавна медијска установа "Радио-телевизија Војводине", Нови Сад, Србија
e-mail: ladislav.lazic@gmail.com

² Мара Скенџић Рељић, новинар, Јавна медијска установа "Радио-телевизија Војводине", Нови Сад, Србија
e-mail: skenzicmara@gmail.com

³ Милијана Кочић, новинар, Јавна медијска установа "Радио-телевизија Војводине", Нови Сад, Србија
e-mail: milijana.kocic@rtv.rs

THE ROLE OF ARTIFICIAL INTELLIGENCE IN THE GENERATION OF DISINFORMATION

Abstract: The emergence of artificial intelligence has led to the rapid creation and distribution of disinformation. The tools it uses are so convincing that it is increasingly difficult to distinguish fake news from the truth. The “flood” of fake news distributed via social networks carries the risks of losing trust in the media, science, institutions, destroying the reputation of individuals, as well as a danger to mental health. On the other hand, together with critical thinking, media literacy and ethical principles of people, artificial intelligence is also an ally in the fight against disinformation. Using inductive, deductive, as well as methods of analysis and synthesis, the conclusion is that artificial intelligence is a weapon in human hands with which we can shape the media space in two opposite directions.

Keywords: *artificial intelligence, fake news, disinformation, media literacy, social networks*

УВОД

Вештачка интелигенција као сама појава али и алат све више мења свакодневицу. Готово да нема друштвене или научне области које није дотакла, а самим тим и утицала на њих. Тако је и у новинарству. Путем ње вести се шире као никада пре, али отвара се питање да ли су оне и веродостојне. Наиме, ВИ (AI - Artificial intelligence) може да произведе потпуно нови, оригиналан садржај који је тешко разликовати од стварног. Једно је сигурно, вештачка интелигенција има своју светлу и мрачну страну. Користи се за стварање и ширење лажних вести и дезинформација, али у исто време је и моћан алат у борби против њих.

На питање да ли је свесна да у исто време ствара лажне вести и бори се против њих, вештачка интелигенција овако одговара:

Не, вештачка интелигенција (ВИ) није свесна у људском смислу. Она не поседује свест, намере, етику или морал. ВИ само извршава задатке за које је програмирана.

Ево како то функционише:

ВИ која креира лажне вести: Ови системи су програмирани да генеришу садржај који је убедљив и реалистичан. Њихов задатак је да, на основу огромних количина података, произведу текстове, слике или видео снимке који испуњавају одређене параметре. Ако је програмирана да створи убедљиву

причу о нечему што се није догодило, она ће то урадити без "сазнања" да је реч о лажима.

ВИ која се бори против лажних вести: Ови системи су програмирани да анализирају податке и препознају обрасце који су типични за дезинформације. Циљ им је да упореде информације са провереним чињеницама, идентификују манипулације на сликама или открију координисане бот мреже.

Дакле, ВИ не осећа никакав унутрашњи сукоб. То је као да користите чекић да закуцате ексер, а затим исти тај чекић да извучете ексер. Чекић не зна да га користите у супротне сврхе, он само извршава задатак.

Одговорност за то како се ВИ користи лежи искључиво на људима који је програмирају и користе. Кључно је поставити етичке стандарде и регулативе како би се спречила злоупотреба ВИ за ширење дезинформација.

Нема сумње да је вештачка интелигенција оружје, у овом случају медијско, којим се барата у зависности од тога да ли желимо да "предност" дамо лажним вестима и дезинформацијама или борби против њих.

"Мрачна страна" ВИ - ширење лажних вести и дезинформација

Заједно са новим технологијама, вештачка интелигенција муњевице ствара и шити лажне вести дезинформације. Најпре се креира садржај, а за то се као алат користе велики језички модели - LLM (Large language models). Они су покретачи генеративне вештачке интелигенције. Захваљујући огромним количинама текстуалних података на којима су тренирани, могу да разумеју и одговарају на питања на природан, људски начин. Користе за креирање текстуалног и визуелног садржаја, прављење сажетака, па чак и писање програмског кода (Salesforce, 2025). Овакви текстови ретко имају граматичке грешке, могу да "пишу пером" од стручњака до просечног корисника, па су зато прилично уверљиви.

Сем текстова, лажни могу бити слике и видео записи и то уз помоћ дипфејке (deepfake) технологије. Она је врста вештачке интелигенције која се користи за креирање убедљивих лажних слика, видео записа и аудио снимака. Дипфејкови често трансформишу постојећи изворни садржај где се једна особа замењује другом. Такође, креирају потпуно оригиналан садржај где је неко представљен како ради или говори нешто што није урадио нити рекао. Највећа опасност коју представљају дипфејкови је њихова способност ширења лажних информација које изгледају као да долазе из поузданих извора. Иако дипфејкови представљају озбиљне претње, они такође имају легитимну употребу, као што су аудио и забава у видео играма, као и апликације за корисничку подршку и одговор на позиве (Techtarget, 2025).

Када се лажни садржај направи, следи његова дистрибуција и то на начин да се што више увећа његов домашај и утицај.

Друштвене мреже Facebook, X (бивши Twitter), Instagram и TikTok су главне платформе за ширење дезинформација лажних вести. На Тиктоку и Инстаграму, лажни видео снимци се муњевито шире, а корисници их често деле не знајући да су лажни и тако несвесно упадају у мрежу лажи.

Садржај деле и шире аутоматизовани налози на друштвеним мрежама које називамо ботовима. На који начин? Веома лако и једноставно. Ботови користе вештачку интелигенцију за масовно дељење, лајковање и коментарисање садржаја и тако стварају варку и илузију да је он популаран и широко прихваћен. Обично није довољно само дељење, већ се примењује и циљано пласирање или тзв. таргетинг (Targeting). Наиме, алгоритми вештачке интелигенције анализирају податке о корисницима, као што су њихови интереси, политичка опредељења и навике претраживања, како би им директно пласирали лажне вести које ће највероватније прихватити (Taboola, 2025). У исто време вештачка интелигенција може да мења садржај и понашање ботова како би се избегла детекција од стране платформе. Тако се на пример лажна вест може изменити на више начина, што отежава препознавање.

На крају, апликације за размену порука WhatsApp и Telegram такође су погодне за ширење дезинформација јер је због енд-то-енд енкрипције тешко пратити порекло лажних порука.

Осим друштвених мрежа и апликација, вештачка интелигенције лажне садржаје пласира и на сајтовима који су веома слични правим информацио-ним порталима. Њихове адресе су сличне адресама познатих медија што додатно збуњује кориснике. "Тамна страна" ВИ може да буде и у служби држава, посебно суперсила попут Русије, Кине и САД. Ова пропаганда дезинформација служи за мешање у унутрашње прилике других земаља.

Примери лажних вести које је креирала вештачка интелигенција

Један од првих и најпознатијих примера примера ураћен дипфејк технологијом била је слика Папе Фрање у белој "пуфер" јакни, која је постала вирална, иако је била у потпуности генерисана вештачком интелигенцијом (Cnn.com, 2023). Манипулацијом уз помоћ вештачке интелигенције није била поштеђена ни краљевска породица у Великој Британији. Наиме, слике принца Вилијама и принца Харија на којима се два брата наводно емотивно грле на крунисању краља Чарлса Трећег, су биле генерисане АИ-јем, и масовно су дељене (Skynews.com, 2023).

Познати примери су и снимци Володимира Зеленског, председника Украјине који се наводно предао (Npr.org, 2022), као и Марка Цукеберга, влас-

ника Фејсбука који је говорио о доманцији светом (Cbs.ca, 2019). Такође, вештачка интелигенција је обрадила и слику ухапшеног америчког председника Доналда Трампа (Bbc.com, 2023).

У тренутно веома актуелном војном сукобу Украјине и Русије понекад се користе алати вештачке интелигенције за прављење лажних видео снимака војника, измишљених изјава званичника, и лажних фотографија разарања које заправо потичу из других конфликта или су потпуно генерисане (Stopfake.com, 2024).

Када је реч о аудио дипфејк снимцима позната је гласовна порука која је имитирала глас бившег председника Сједињених Америчких Држава Џоа Бајдена, којом је позивао гласаче да не изађу на изборе у Њу Хемпширу, а која је креирана путем вештачке интелигенције (Bbc.com, 2024). У Словачкој за време избора 2023. године проширио се аудио дипфејк снимак који је наводно приказивао политичара како разговара о намештању избора, што је изазвало значајне контроверзе (Ipimedia.com, 2023).

Лажне вести могу да изазову поремећаје и у економији. Тако је вештачком интелигенцијом обрађена зграда у близини Пентагона у Вашингтону која је наводно горела, што је прозороковало краткотрајан, али значајан поремећај на америчкој берзи (Cnn.com, 2023). Такође, сличан пример је и ширење лажног извештаја да је SEC, Комисија за хартије од вредности Сједињених Америчких Држава одобрила ETF (exchange-traded fund) спот-листиран за биткоин, што је изазвало промену цена биткоина (Reuters.com, 2024). У Хонг Конгу је банкарски службеник био преварен видео конференцијом у којој су алати вештачке интелигенције искоришћени за прављење лица познатих раднику, па је верујући у то што види, на захтев једног од директора на рачун пребацио више од 25 милиона долара (Cnn.com, 2024).

Вештачка интелигенција и њени алати користе се за креирање веродостојних, али попуно измишљених цитата политичара и званичника. Поједини сајтови почели су да је користе за писање вести, а аутори који су потписани као ствараоци вести нису стварни (Cnn.com, 2024). Интересантан је пример новинара Бена Блека, који сваке године на Светски дан шале објављује лажне приче. Он је 2020. године написао чланак како је Кумбран (Cwmbran), градић у Јужном Велсу ушао у Гинисову књигу рекорда због највише кружних токова по квадратном километру. Међутим, иако је ту вест истог дана преправио, вештачка интелигенција је пет година касније ову информацију представила као тачну (Bbc.com 2025).

ВИ је коришћена и за стварање лажних слика и снимака спашавања у подручјима погођеним природним катастрофама, што отежава спасиоцима

да филтрирају информације и фокусирају се на стварна места и локације где је помоћ заиста потребна (Cbsnews.com, 2024).

Ризици које стварају лажне вести

После стварања и ширења лажних вести и дезинформација и њихове “конзумације” настају ризици. Када је тешко препознати шта је лаж, а шта истина, губи се поверење у медије, науку, институције, што може да доведе до масовног скептицизма.

Лажним снимцима и аудио записима врло је лако дискредитовати угледне личности из свих сфера друштвеног живота, али и обичне грађане. Када се углед наруши, па макар и на овај подмукао начин, често га је тешко поново изградити.

Овај начин дискредитације може да доведе до манипулације јавним мњењем. Наиме, у изборним кампањама циљ је да се дезинформацијама и лажним вестима изгуби поверење бирача у политичке противнике. Ове активности могу да изазову поделе, поларизацију друштва и политичку нестабилност.

У исто време дипфејк технологије идеалне су за новчане преваре, а ширење лажних вести о компанијама може да доведе до финансијских губитака.

На крају сви ови ризици када се саберу представљају опасност по ментално здравље људи. Константна изложеност лажним вестима, дезинформацијама, дипфејк садржајима изазива неповерење, психички замор, анксиозност, дефетизам...

“Светла страна” ВИ - борба против лажних вести и дезинформација

Као што их и ствара и шири, тако се вештачка интелигенција користи и у борби против лажних вести и дезинформација. Постоји више начина како да то учинимо.

За откривање лажних вести, користе се ВИ алгоритми који могу да анализирају велике количине текста и тако препознају обрасце типичне за лажне вести. То укључује анализу тона, коришћење сензационалистичких наслова и препознавање наратива који су уобичајени у теоријама завере. Тако на пример Модели за обраду природног језика (NLP, Natural language processing) могу да упореде текст са провереним изворима и оцене његову веродостојност (101Blockchains, 2024).

Осим лажних вести, ВИ може да открије и ботове. Наиме, њени алгоритми анализирају активности као што су брзо дељење садржаја, генеричке профилне слике, координисање коментара...Када их открију ствара се могућност да се изолују и склоне са друштвених мрежа.

Алати ВИ се користе и у анализи видео и аудио снимака како би открили дипфејк манипулације. Алгоритми траже суптилне неправилности које људско око не може лако приметити, попут неприродних покрета лица, осветљења и сенки, или несавршености у звуку. Ови алати могу да препознају да ли је нечији говор или изглед вештачки генерисан или измењен.

У повоју је и технологија која ставља “водене жигове” на дигитални садржај генерисан ВИ. То у пракси представља означавање садржаја који је креиран вештачком интелигенцијом, па корисници лако препознају његово порекло (Datacamp.com, 2025).

Системи засновани на ВИ могу да анализирају вести, објаве на друштвеним мрежама, аутоматски проверавајући и упоређујући тврдње са поузданим базама података у реалном времену.

Такође, алати ВИ служе новинарима за проверу чињеница. Уместо да “пешке” проверавају тврдње, користе их како би брзо добили почетну анализу и идентификовали најсумњивије информације.

Ипак, да би борба против лажних вести и дезинформација била потпуна, сем технологије неопходно је и људско искуство и запажање. Ту пре свега мислимо на проверу извора и аутора вести, урл адреса...Лажне вести најчешће немају аутора, а и ако га имају брзом претрагом на интернету лако је установити да ли та особа заиста постоји, односно да ли је новинар. Сајтови са којих се дезинформације пласирају обично личе на праве портале, али имају грешке у урл адресама.

Како би биле што читаније лажне вести често имају бомбастичне и драматичне наслове, што одмах баца сумњу на њих, па се тако лако примете. Такви текстови, иако написани алатима ВИ нису решили проблем суптилности језика. Тако се примећује понављање фраза, необичне конструкције реченица и логичке грешке. Такође, ови текстови не садрже имена извора или их нејасно наводе, као ни конкретне податке и статистике. Укратко не позивају се на проверљиве изворе као поуздане и тачне вести. Код дипфејк снимака људско око на пример лако може да примети нереално осветљење, неприродан покрет ока, неправилне сенке на лицу као и усклађеност звука са уснама...

И поред свих начина и напора да се уђе у траг лажним вестима и дезинформацијама тешко је очекивати да ћемо их препознати и разоткрити без едукације и медијске писмености. Тим пре што се иста технологија користи за њихово прављење и откривање. Та “трка у наоружању” ће бити константна, па је веома важно едуковати и медијски описменити новинаре, али и обичне интернет кориснике. Медијска писменост подстиче кориснике да не прихватају информације здраво за готово, већ да их преиспитују. Она је неопходна

како би сами могли да процененимо кредибилитет извора и да схватимо да је дезинформација велика претња по стабилност друштва.

Критичко размишљање, намера и етика свакако су основе провере сваке вести. Без јаких етичких правила и препознавања манипулација технологија није довољна у борби против чажних вести и дезинформација. Због тога пре него што поделимо вест треба да се запитамо да ли је извор поуздан, односно да ли исту информацију можемо да пронађемо на неком другом месту за које проверено знам да је поуздано. Такође, ваља размислити и о томе да ли нам је вест превише добра или лоша да би била истинита. И то не само нама, већ и онима којима желимо да је поделимо.

У ширење медијске писмености неопходно је уљкучити новинарску струку, невладине, организације, образовне установе како би спровели програме обуке. Веома је важно показати и објаснити психолошке тактике (изазивање страха, поларизација, наметање кривице...) које се користе у лажним вестима.

ЗАКЉУЧАК

Коришћење вештачке интелигенције доноси одређене ризике и опасности. Није немогуће да ће једног дана постати сувише напредна и способна да сама мења циљеве. Ако се "отме контроли" могла би да ради против интереса човечанства што смо имали прилику да видимо у појединим научно фантастичним филмовима. Такође, претерано ослањање на њу чини људе лењим и зависним, тим пре ако она доноси одлуке или изборе уместо нас. Због тога је важно регулисати правне оквире и етичке принципе на основу којих би се одговорно користила. У циљу борбе против лажних вести и дезинформација и друштвене мреже и медијске платформе треба да појачају своје напоре, користећи ВИ за брже препознавање и уклањање лажног садржаја. Са друге стране, владе и међународне организације дужне су да донесу и спроводе законску регулативу која би дефинисала одговорност и казне за ширење дезинформација. Извесно је да ће борба против стварања и ширења лажних вести и дезинформација уз помоћ вештачке интелигенције бити непрестана, али и вишеслојна. Технологија сама по себи неће бити довољна, јер иако напредна понаша се као двосекли мач. Колико год била моћна у против њих, толико је снажна и у стварању истих. Напослетку, технологију контролишу људи који је користе на добар или лош начин. Без људског фактора, а ту подразумевамо критичко размишљање, јаке етичке принципе и општу ме-

дијску писменост, борба против лажних вести неће дати праве резултате. Без медијске писмености, чак и најбоља технологија је бескорисна.

ИНТЕРНЕТ ИЗВОРИ

1. What are LLMs (Large Language Models)? Available at: <https://www.salesforce.com/artificial-intelligence/what-are-large-language-models/> (Accessed: September 2, 2025).
2. What is deepfake technology? Available at: <https://www.techtarget.com/whatis/definition/deepfake> (Accessed: September 2, 2025).
3. AI audience targeting: Everything you need to know, Available at: <https://www.taboola.com/marketing-hub/ai-audience-targeting/> (Accessed: September 2, 2025).
4. Look of the Week: What Pope Francis' AI puffer coat says about the future of fashion, Available at: <https://edition.cnn.com/style/article/pope-francis-puffer-coat-ai-fashion-lotw> (Accessed: September 3 2025).
5. Fake AI images keep going viral - here are eight that have caught people out, Available at: <https://news.sky.com/story/fake-ai-images-keep-going-viral-here-are-eight-that-have-caught-people-out-13028547> (Accessed: September 3, 2025).
6. Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn, Available at: <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> (Accessed: September 3, 2025.)
7. What a deepfake video of Mark Zuckerberg reveals about how we're manipulated online, Available at: <https://www.cbc.ca/radio/day6/raptors-dreams-a-zuckerberg-deep-fake-bruce-cockburn-on-mmivg-video-game-streaming-summer-reads-and-more-1.5174349/what-a-deepfake-video-of-mark-zuckerberg-reveals-about-how-we-re-manipulated-online-1.5174405> (Accessed: September 3, 2025.)
8. Fake Trump arrest photos: How to spot an AI-generated image, Available at: <https://www.bbc.com/news/world-us-canada-65069316> (Accessed: September 3, 2025).
9. How AI-Generated Content Distorts Truth About War and Russian War Crimes: Detecting and Analyzing AI-Generated Images, Available at: <https://www.stopfake.org/en/how-ai-generated-content-blurs-the-truth-about-war-and-russian-war-crimes-detecting-and-analyzing-ai-generated-images/> (Accessed: September 3, 2025).
10. Fake Biden robocall tells voters to skip New Hampshire primary election, Available at: <https://www.bbc.com/news/world-us-canada-68064247> (Accessed: September 3, 2025).

11. Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse, Available at: <https://ipi.media/slovakia-deepfake-audio-of-dennik-n-journalist-offers-worrying-example-of-ai-abuse/> (Accessed: September 3, 2025).
12. Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion, Available at: <https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion> (Accessed: September 3, 2025).
13. SEC probing fake post on its X account, bitcoin ETFs not yet approved, Available at: <https://www.reuters.com/technology/us-sec-has-not-approved-bitcoin-etfs-social-media-account-compromised-2024-01-09/>, Accessed: September 3, 2025).
14. Finance worker pays out \$25 million after video call with deepfake 'chief financial officer', Available at: <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk> (Accessed: September 3, 2025).
15. A national network of local news sites is publishing AI-written articles under fake bylines. Experts are raising alarm, Available at: https://www-cnn-com.translate.google.com/2024/05/30/media/ai-bylines-local-news-hoodline?_x_tr_sl=en&_x_tr_tl=sr&_x_tr_hl=sr&_x_tr_pto=sc, (Accessed: September 3, 2025).
16. Bogus videos from the hurricanes are going viral. Here's how to spot old and fabricated footage online, Available at: <https://www.cbsnews.com/news/hurricane-viral-video-how-to-spot-old-fabricated-ai-footage/> (Accessed: September 3, 2025).
17. Google AI presented my April Fools' story as real news, Available at: <https://www.bbc.com/news/articles/cly12egq5ko>(Accessed September 3, 2025).
18. Best Natural Language Processing (NLP) Models, Available at: <https://101blockchains.com/best-nlp-models/> (Accessed: September 4, 2025).
19. AI Watermarking: How It Works, Applications, Challenges, Available at: <https://www.datacamp.com/blog/ai-watermarking> (Accessed: September 4, 2025).

Цитирање – АПА стил:

Лазић, Л., Скенцић Рељић, М., Кочић, М. (2025). Вештачка интелигенција и ширење дезинформација. У: Б. Санчанин и А. Пењишевић (ур.), *Зборник радова VII међународне научне конференције Вештачка интелигенција у креирању медијских садржаја* (стр. 162-172). Сремски Карловци: Центар за културу, едукацију и медије; Источно Сарајево–Соколац: Висока школа за услужни бизнис.

Рад је примљен: 8. септембра 2025.

Received: September 8, 2025

Прихваћен за објављивање: 2. октобра 2025.

Accepted: October 2, 2025