

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ БЕЗБЕДНОСТИ
КАТЕДРА СТУДИЈА БЕЗБЕДНОСТИ



**ШИРЕЊЕ ЛАЖНИХ ВЕСТИ НА ДРУШТВЕНИМ
МРЕЖАМА**

- ДИПЛОМСКИ РАД -

Ментор:

Др Ана Ковачевић
Ванредни професор

Студент:

Матеја Петковић
74/17

Београд, 2022.

САДРЖАЈ

1. Увод	1
2. Лажне вести	3
2.1. Дефинисање појма лажних вести	3
2.2. Историјат лажних вести	4
2.3. Савремени ефекти лажних вести	6
3. Друштвене мреже	9
3.1. Историја развоја друштвених мрежа на интернету и њиховог утицаја на живот човека	9
3.2. Лажне вести на друштвеним мрежама	13
4. Потрага за решењем проблема лажних вести	17
4.1. <i>ClaimBuster</i> – Систем за откривање и оспоравање обманујућих чињеница	18
4.1.1. Преглед система <i>ClaimBuster</i>	19
4.2. Комбиновање људског и алгоритамског приступа у борби са лажним вестима	22
5. Закључак	24
Литература	26

1. Увод

У овом раду се истражује проблем лажних вести, са акцентом на феномен ширења лажних вести на друштвеним мрежама. За проблем се може рећи да је веома актуелан и широког обухвата, те га није лако у потпуности сагледати, до краја анализирати и највише од свега, решити. У савременом свету лажне вести и други сродни типови дезинформација постају све значајнији проблем. Циљ рада је да проблем представи, али и да представи неке од идеја како се борити са проблемом, ублажити га или савладати. За почетак потребно је извршити дефинисање лажних вести, што је први задатак са којим се рад суочава. Дефинисање се врши помоћу доступне академске литературе, у којој постоје различите варијанте дефиниција, у зависности од приступа, мада им је срж заједничка. Потом се проблем лажних вести контекстуализује, кратким историјским прегледом и представљањем размера и актуелних показних типова овог проблема. Затим се сличним приступом представљају и друштвене мреже. Након дефинисања, историјског прегледа и посматрања утицаја који друштвене мреже имају на живот савременог човега посматра се на који се начин лажне вести шире на друштвеним мрежама. Овде се стиче јаснија слика о значају проблема којим се рад бави, његовим ефектима, размерама и донекле последицама. Остатак рада посвећен је напорима који су последњих година улагани да се проблем ширења лажних вести на друштвеним мрежама ублажи. Идеја је да се изврши анализа могућности изналажења софтверских решења за овај проблем, односно колико су се до сада успешним показали напори те врсте, као и посматрање потенцијала софтверских решења у будућности. Представљен је ClaimBuster систем за изналажење и проверу сумњивих тврдњи онлајн и на телевизији. Овај систем замишљен је тако да врши потпуну проверу, с краја на крај, где сам уочава тврдње које вреди проверити, затим их истражује, проверава и нуди их заинтересованом кориснику уз контекст из ког је тврдња извучена. Затим је у раду

извршена кратка анализа резултата и могућности проналажења софтверских решења за проблем ширења лажних вести. Додатна пажња у раду посвећена је етичкој страни проблема лажних вести, односно етички проблем решавања наведеног проблема, где се јављају опасности укидања слободе говора и цензурисања.

2. Лажне вести

2.1. Дефинисање појма лажних вести

Лажне вести постају један од најзначајнијих проблема у савременом друштву. Улога медија у животу просечног човека се проширила, информације и разни садржаји се деле у огромним количинама невероватном брзином. Човек не може да постигне да „профилтрира“ сваку вест коју добије и презасићеност информацијама олакшава и лажним вестима да се шире у већим количинама него икада раније. Лажне вести се могу дефинисати као нетачне или обмањујуће информације које су представљене као вести (Tufekci, 2018). Међутим, дефиниција није утврђена и појам лажних вести за различите ауторе може имати различито значење. Како у академској заједници, тако и у свакодневним разговорима постоје разне интерпретације тога шта су лажне вести и шта све спада у лажне вести. Уколико се фокус постави на академску заједницу наилазиће се на радове који садрже различите типологије, односно разврставања типова лажних вести у покушају да се оне ближе одреде. Један истраживачки подухват који се бавио дефинисањем лажних вести проучавањем 34 различита научна рада из ове области довео је до диференцирања следећих типова лажних вести: сатире, пародије, фабриковања, манипулације, оглашавања и пропаганде (Tandoc et al., 2017). Наведена типологија заснована је на две димензије које варирају од типа лажних вести: на фактичности и намеру да се обмане. Ово може значајно да послужи у разумевању значења појма лажних вести, али други аутори су истакли потребу да се појам додатно операционализује како би се лажне вести лакше откривале у будућности, коришћењем алгоритамских метода (Molina et al., 2019). На овом задатку наилази се на доста препрека, не само због комплексности самог појма и његове манифестације, већ и због тога што се данас често употребљава и у политичке сврхе, како би се одређени новински извори или медијске куће дискредитовали. Наведени аутори створили су таксономију од осам категорија

различитог онлајн садржаја са сврхом да омогући алгоритамски засновано разоктривање лажних вести. Ту спадају: праве вести, лажне вести, поларизован садржај, сатира, погрешно извештавање, коментарисање, убеђивање и грађанско новинарство (Molina et al., 2019).

2.2. Историјат лажних вести

Лажне вести нису нов феномен. Људи су одувек лажирали информације, али се форма и начин преношења јесу значајно мењали кроз векове. Гутенбергов изум штампарије 1493. довео је до неупоредиво веће количине преноса информација, а самим тим и дезинформација, што нас неколико векова касније доводи до прве велике новинске преваре. Реч је о шест артикала које је 1835. године тадашњи њујоршки лист „Д Сан“ објавио на тему наводног открића живота и цивилизације на Месецу (видети Сliku 1). Догађај је остао упамћен као прва велика лажна вест (Thornton, 2009). Значајан искорак у развоју лажних вести догодио се током Другог бурског рата (1899-1902), када је у Великој Британији створен бурски стереотип. Британска армија је преко тадашњих новина и служећи се додатно и памфлетима и плакатима створила тај стереотип како би убедила британску јавност да подржи иначе непопуларну одлуку да се поведе рат (Kent, 2013). Развој комуникације једног према многима проналаском радио комуникације и телевизије у 20. веку обележила је 20. век (Posetti & Matthews, 2018). Потом се у 20. веку показало колико пропаганда може бити моћно оружје са огромним последицама. Два светска рата и Хладни рат, с једне стране и развој медија као што су радио, телевизија и напослетку интернет, с друге стране довели су до различитих начина (зло)употребе информација које су умногоме обликовали људске животе.



Слика 1: Илустрација површине Месеца објављена у новинама Д Сан(Извор: Wikipedia, 2022)

Развој интернета крајем 20. века и друштвених мрежа у 21. веку драматично су увећали ризик од ширења лажних вести, дезинформација, пропаганде и превара. Данас се на друштвеним мрежама путем комуникације многих са многима и погрешан и обмањујући садржај масовно дели, те постаје виралан. Овакво стање карактерише се као кризно и прети да у потпуности „угуши“ новинарство (Posetti & Matthews, 2018). Појам и истраживање феномена лажних вести везује се за средину друге деценије 21. века, тачније за период током и након кампање председничких избора у Америци 2016. године. Од тада термин „fake news“ улази у фокус интересовања медијских и истраживача у пољу информационих технологија, али и шире јавности као конзумента медијског садржаја (Јовановић и Богдановић, 2018). Оксфордски речник је исте године прогласио реч пост-истина за реч године (Oxford languages, 2016). Реч је описана као придев који занемарује објективне чињенице у обликовању јавног мњења. Пост-истина је друштвени феномен где људи лакше прихватају информације које се ослањају на њихова уверења и веровања, а не на доказе и чињенице. Пост-истина је срж лажних вести (Груловић, 2021). Данас се одвија процес све веће поларизације у друштву о чему сведоче екстремно супротстављени ставови исказани на

интернету, путем друштвених мрежа. Иако су релативно мирна времена у току (у односу на крвави минули век), стиче се утисак да се константно одвијају сукоби неке врсте на друштвеним мрежама. Једно од таквих поља за дискусије без правила јесте друштвена мрежа Твитер. Једноставно и бесплатно коришћење омогућава људима да с лакоћом износе ставове и мишљења за широку публику коју ограничава само непрегледно море објава. У таквим условима често се истичу они „најбучнији“ и „најекстремнији“. Људи потом прибегавају свакаким опцијама да докажу и поткрепе своје ставове, а најчешће су то информације које спадају у неку од категорија лажних вести. Истина је увек „негде између“.

2.3. Савремени ефекти лажних вести

Једна област која је нарочито погођена лажним вестима јесте здравство/медицина. Увелико су познате вести о алтернативној медицини, као и теорије о „фармацеутској мафији“ или о вакцинама које изазивају аутизам. Ово већ „ужарено поље“ вишеструко се проширило од како је избила пандемија Ковид-19 вируса, што је довело (и доводи) до озбиљних здравствених и политичких последица (Carrion-Alvarez & Tijerina-Salina, 2020).

У току писања овог рада још једна криза потресла је свет, а то је рат у Украјини. Ова криза још једном у низу показала је значај и размеру проблема лажних вести. Рат је нарочито током првих недеља изазвао праву бујицу информација на друштвеним мрежама, а самим тим и дезинформација, које су се делиле претежно на Твитеру, Фејсбуку и ТикТоку. Количина дезинформација је створила неку врсту дигиталне ратне магле. Неосноване информације које објављују Владе Украјине и Русије, се потом масовно деле на друштвеним мрежама.

Експерти који се баве дезинформацијама на друштвеним мрежама истичу да је ова криза одвела проблем лажних вести корак даље у односу на пандемију Ковида-19 и других прошлих криза. Главни разлози за то јесу сам садржај који се

дели и природа ситуације која укључује супротстављене стране у конфликту. Садржај чине претежно слике и снимци, за разлику од пандемије Ковида када су лажне информације биле присутне у текстуалним садржајима. Проблем је што је слике и снимке теже проверити, како аутоматским системима, тако и професионалцима. За слике могуће је користити Гугл претраживач за слике, тако да то делимично олакшава посао, док је за проверу снимака другачији приступ. Најпре се проверава да ли упитни снимак већ постоји онлајн, односно да ли је већ раније објављиван. Ако се утврди да није и да је снимак актуелан, онда се приступа разним аналитичким алатима како би се утврдила веродостојност садржаја. На пример, анализира се положај сенки или се терен упоређује са сателитским снимцима да би се утврдило да ли је представљена локација аутентична.

Чињеница да се води рат сведочи о двеју страна са екстремно супротстављеним интересима, што се наравно преноси и у сферу информација. Пропаганда је данас саставни део сваког сукоба, тако да је проблем са лажним вестима у овом случају нарочито закомпликован. На све то, ситуација је таква да изазива јаке емотивне реакције код људи корисника друштвених мрежа, што ослабљује способност да се рационално испита извор неке информације, већ се брже приступа дељењу садржаја у жељи да се на неки начин „помогне“ (Duffy & Metz, 2022).

Анализе показују и да лажне вести сада брже прелазе државне границе него што је био случај током пандемије Ковид вируса (Duffy & Metz, 2022). Објашњење ове појаве највероватније је садржено поново у врсти садржаја који се дели; видео садржаји превазилазе језичку баријеру за разлику од текстуалних садржаја. Проблем је нарочито у масовности дељених информација, јер када корисник друштвених мрежа бива изложен разним непровереним снимцима и сликама

може на основу неких погрешних идеја и уверења формирати своје ставове и уопште поглед на тренутна дешавања у свету (Duffy & Metz, 2022).

Један од примера лажних вести је прича о „Духу из Кијева“, украјинском пилоту који је оборио многе руске авионе. Негде је реч о шест, а негде о чак 49 авиона. Појавиле су се чак и наводне фотографије на којима се налази пилот. Ова прича је стекла огромну популарност на друштвеним мрежама, масовно се делила, да би се потом испоставила као нетачна. Нигде нису постојале званичне информације или доказ о једном пилоту који је оборио више руских авиона. Поменуте фотографије су фотографије које су објављене од стране Министарства спољних послова Украјине 2019. године, а на којима се налази адвокат из Аргентине (Ташковић, 2022). Због популарности ове приче, али донекле и њене наивности, изазвала је многе реакције. Након што је утврђено да је прича нетачна, појавило се доста „шљивих“ слика које представљају разне јавне личности „Духом из Кијева“. Ове намерно очигледно сковане дезинформације стварају додатну конфузију на друштвеним мрежама и развијају дигиталну ратну маглу која од самог почетка карактерише овај сукоб на истоку Европе.

Фокус овог рада је на ширењу лажних вести и могућностима да се томе стане на пут на засигурно најзначајнијем пољу њиховог ширења – друштвеним мрежама. Због тога следи кратак преглед значаја и ефекта које саме друштвене мреже имају на живот савременог човека, а потом и на који начин утичу на проблем ширења лажних вести.

3. Друштвене мреже

Да би било могуће проучавање утицаја и ширења лажних вести, а све у сврху њиховог сузбијања, неопходно је најпре осмотрити "терен" на ком се оне масовно шире, а то су друштвене мреже. Због тога у раду следи кратак опис друштвених мрежа, њиховог настанка и утицаја које оне имају на животе људи.

Друштвене или социјалне мреже су онлајн веб сервиси који корисницима омогућавају разноврсне видове комуникације и могућност личне презентације. Неке од најпопуларнијих друштвених мрежа које су занимвиље за већину корисника са наших простора су Фејсбук, Инстаграм, Линкедин и др. Присутност на већини друштвених мрежа је бесплатан(Крстић, 2018).

Чини се да данас готово нема човека који живи и ради у савременом свету, а да није корисник барем неке од највећих друштвених мрежа. Према истраживањима које спроводи Statista.com неку форму друштвених мрежа користи 4,2 милијарде људи на свету. На глобалном нивоу, 53.6 одсто светске популације је на друштвеним мрежама, а гледано по регионима Западна и Северна Европа предњаче, са по 79 одсто популације која користи неку од друштвених мрежа. Највећа платформа је Фејсбук, који броји 2,9 милијарди корисника широм света. Људи масовно приступају друштвеним мрежама са својих мобилних уређаја, та групација броји 4,15 милијарди. Просечно време проведено на друштвеним мрежама од стране корисника интернета широм света на дневној бази износи 145 минута. Ово су неки од статистичких података који осликавају масовност друштвених мрежа (Dixon, 2022).

3.1. Историја развоја друштвених мрежа на интернету и њиховог утицаја на живот човека

Први сајт који се по дефиницији може посматрати као друштвена мрежа је Sixdegrees.com, који је настао 1997. године. Корисници су могли да постављају

профилне фотографије, упознају пријатеље, шире своје друштвене групе, успостављају конекције и да шаљу поруке. Пред почетак новог миленијума, почели су да се појављују сајтови за блогове међу којима и LiveJournal који је подстицао људе на међусобно „праћење“ и прављење група са међусобном интеракцијом. Потом 2000. године настаје први сајт друштвене мреже који се финансирао из реклама, LunarStorm. Сајт је настао у Шведској и био је намењен пре свега тинејџерима. До 2007. је накупио 1,2 милиона чланова, да би се затворио 2010.

Прва већа друштвена мрежа звала се Френдстер (Friendster) и настала је 2002. године. Користила је за све уобичајене ствари, али такође и као сајт за упознавање партнера и за откривање догађаја, хобија и слично. 2011. године на Френдстер су уведене друштвене игрице. Угашен је 2015. Линкдин (LinkedIn), прва позната друштвена мрежа за бизнис настала је 2003. године, а користи се и данас и има скоро 800 милиона корисника. Исте године настао је Мајспејс (MySpace) који је убрзо затим постао главна друштвена мрежа свог времена. Омогућио је корисницима да у потпуности прилагоде свој профил, као и да објављују музику и видео записе. Мајспејс је прилично револуционаран за своје доба и представља значајан део историје друштвених мрежа (McFadden, 2020).

Прва, затворена верзија Фејсбука настала је 2004. године, да би две године касније била развијена и верзија за тржиште. Фејсбук је и дан данас једна од најпопуларнијих платформи друштвеног умрежавања на интернету (а самим тим и на свету) и како је у раду већ споменуто, има највише корисника од свих друштвених мрежа. Од 2006. до данас бележи запањујућ раст и представља један од најпосећенијих сајтова на интернету уопште. У међувремену је 2005. године настао и Јутјуб, такође веома значајна друштвена мрежа која омогућава корисницима да деле видео садржај са осталим корисницима и који и данас има огроман број корисника. 2006. године настаје Твитер који се неизоставно узима у

обзир када се разговара о најзначајнијим друштвеним мрежама, а који омогућава дељење кратких текстуалних садржаја. Популаран је као извор информација, нових вести и зато што корисницима даје могућност да комуницирају са познатим личностима.

Година када су многа предузећа заиста препознала маркетиншки потенцијал друштвених мрежа је 2008. Од тада су настали многи бизнис профили на друштвеним мрежама. Појава Инстаграма 2010. је следећи важан корак у развоју друштвених мрежа, јер је ова платформа за дељење фотографија и видео снимака убрзо постала једна од највећих друштвених мрежа на свету. Само у првих неколико месеци Инстаграм је прибавио 2 милиона корисника. Фејсбук је увидео потенцијал Инстаграма и купио га је 2012. године и од тада до данас у константном је расту. Данас се нарочито користи у сврхе пословања, као рекламна мрежа где се промовишу људи и предузећа. У наредним годинама развијено је мноштво друштвених мрежа са разним циљевима и сврхама. Значајно је поменути још и ТикТок који је кинеска апликација намењена за западно тржиште. Ова платформа је настала 2016. године и од тада је стекла огромну популарност, а пре свега је намењена за дељење забавног садржаја (видети Сliku 2) (McFadden, 2020).



Слика 2: Најпопуларније друштвене мреже(Извор: CNN, 2022)

Могуће је дакле закључити да су се друштвене мреже у великој мери развијале у протекле две деценије, упоредо са развојем технологије и улоге коју интернет има у свакодневном животу једног човека. Друштвене мреже привлаче пуно пажње будући да чине саставни део свакодневнице просечног човека. Као такве оне имају велики утицај на живот корисника, а тај утицај је сложен и карактеришу га како позитивне, тако и негативне одлике. Предности друштвених мрежа су махом јасне и познате; оне су истакнуте од стране самих платформи, мада се некад „подвуче“ и нека неочекивана предност и открије тек након одређеног времена. Неке од главних предности друштвених мрежа свакако су: лакоћа одржавања контакта са породицом и пријатељима, могућност повезивања људи по струци, хобијима, интересовањима, уверењима и др., затим могућност упознавања нових људи, бројне предности које присуство на друштвеним мрежама доноси предузећима, уопште лакоћа приступа „главним“ друштвеним мрежама итд. Међутим, последњих година се све више расправља и о одређеним манама и недостацима друштвених мрежа. Неки од недостатака (а који потичу од начина

на који људи користе друштвене мреже) јесу: лажна представа о стварности коју друштвене мреже могу да изазову, када површне интеракције путем друштвених мрежа постану нека врста „замене за живот“, губљење времена на друштвеним мрежама, чињеница да друштвене мреже стварају зависност и проблеми са самопоуздањем који настају (Dalomba, 2020).

С обзиром на то колико сложен утицај друштвене мреже врше на човека, неки истраживачи предности и недостатке посматрају тако што издвајају одређене области или раздобља људског живота и посматрају ефекат друштвених мрежа на њих. Тако на пример посматрајући утицај друштвених мрежа на децу и тинејџере можемо закључити да им оне олакшавају развој социјалних способности, инспиришу их и помажу им да развијају способности читања и писања (ако имају позитивне узоре), док са друге стране постоји опасност да им створе лажну слику о изгледу људског тела, да их уведе у ризично сексуално понашање и да им поквари успех у школи (Akram & Kuram, 2017). Истраживање које се бавило ефектом друштвених мрежа на студенте дошло је до резултата који указују на негативан ефекат на академски успех као последицу одузимања времена и ометања приликом предавања и учења. Са друге стране, позитиван ефекат је на однос између студената (Wang et al., 2011). Консензус је да су друштвене мреже алат, који може бити добро или лоше коришћен од стране човека. Истиче се важност едукације, нарочито младих људи и деце како би извукли оно најбоље што друштвене мреже могу да понуде и да би избегли разне „замке“ у које могу „упасти“.

3.2. Лажне вести на друштвеним мрежама

Једна несумњиво негативна појава на друштвеним мрежама јесу лажне вести. Све више људи се окреће према друштвеним мрежама када је реч о

формирању политичких мишљења и уверења. Светски економски форум је 2013. године препознао масовну дигиталну дезинформисаност као значајан технолошки и геополитички ризик(Howell, 2013). Избори одржани у САД 2016. године обележени су расправом о лажним вестима и дезинформацијама, а главно „бојиште“ била је друштвена мрежа Твитер. Једно истраживање спроведено је на бази података коју је сачињавало 171 милион „твитова“ из периода последњих пет месеци пред изборе. Од тога 30 милиона „твитова“ од стране 2,2 милиона корисника садржавало је везу (линк) одређене вести. Истраживање је дошло до закључка да је 25 одсто тих „твитова“ делило лажне или екстремно пристрасне вести. Све ово је имало утицај на исход избора (Bovet & Makse, 2019). У завршна три месеца америчких избора 2016. године 20 најуспешнијих лажних вести у вези са изборима изазвало је више ангажовања (дељења, реакција и коментара) на Фејсбуку него 20 најуспешнијих вести највећих новинарских извора (Њујорк Тајмс, Вошингтон поуст и Ен-би-си њуз). Добар део ових лажних вести настао је и пласиран из Велеса у Македонији, па је и прича о томе како је група Македонаца утицала на изборе у Америци постала нашироко позната (Hughes & Waismel-Manor, 2020).

Утврђено је да постоји 70 одсто више шанси да се лажна вест „ритвитује“ него права (тачна) вест. Такође тачној вести треба око шест пута више времена да доспе до 1500 људи него што је случај са лажном вешћу. Када је реч о непрекинутим низовима „ритвитова“, лажне информације достижу низ од 10 „ритвитова“ око 20 пута брже него чињенице. Аутори овог истраживања сматрају да је узрок овог феномена у људској психологији; они сматрају да људи једноставно воле новитете. Док су вршили истраживање препознали су различите емоционалне профиле код људи који деле лажне вести у односу на оне који деле чињенице. Лажне вести изазивају јаче реакције и због тога се чешће деле (Dizikes, 2018). Постоји и другачији приступ када се тражи узрок феномена масовног дељења лажних вести на друштвеним мрежама. Људи су подложни когнитивним

пристрасностима; људски ум нове информације константно прилагођава и усклађује тако да се оне уклопе са дотадашњим знањем. Свако преферира информације које добија од људи којима верује, најчешће припадницима властите групе. У условима преоптерећености информацијама људи се често ослањају на своје когнитивне пристрасности да бирају чему ће поклонити пажњу. Друштвене мреже повећавају ове пристрасности на штетан начин, тако што кориснике усмеравају и повезују са истомишљеницима. Претраживачи и друштвене мреже располажу огромном количином података о сваком кориснику и на основу њих приоритетно нуде информације са којима ће се корисник највероватније сложити; корисник добија информације које највише одговарају његовим уверењима што доводи да он своје уверење додатно учврсти. Процес се одвија аутоматски и не води рачуна о томе да ли су подаци које корисник добија и дели истинити или лажни. Последица оваквог програмирања јесте поларизација неистомишљеника која води у екстремизам, као и настанак одређених ехо-комора. Дезинформисана ехо-комора се додатно затвара и иде све дубље у своје лажно уверење, делећи унутар себе лажне информације и вести (Menczer & Hills, 2020). Истраживање из 2011. године показало је да су ехо-коморе толико изражене, да је могуће са великом прецизношћу предвидети за кога ће један корисник Твитера гласати, само на основу података који су доступни на тој друштвеној мрежи (Conover et al., 2011).

Важно је поменути и још једну појаву која доводи до масовног умножавања броја лажних вести, а то су лажне вести генерисане од стране вештачке интелигенције. Модели вештачке интелигенције могу да имитирају људски начин писања вести на тај начин да их је скоро немогуће открити као вештачке творевине. Ови модели се користе у сврху обмањивања делова популације или пропаганде и масовно преплављују друштвене мреже (Sawant, 2021). Истраживања показују да појединци углавном нису у стању да разликују текстове писане од стране људи и оних који су стварани уз помоћ вештачке интелигенције (Kreps, et al. 2022). Ово

отвара могућност злоупотребе тако што потенцијал ширења дезинформација расте немерљивом брзином. Није претенциозно тврдити да човек неутемељених мишљења и ставова који црпи информације путем друштвених мрежа, од људи и извора којима верује врло лако може упасти у неку од бројних замки које су представљене у овом поглављу. Такав човек је на друштвеним мрежама константно у опасности да дође у ситуацију да базира своје виђење света на лажима и загарантоване негативне последице тога варирају од бенигних до катастрофалних.

Како утицати на овај масовни проблем и у што већој мери спречити ширење лажних вести на друштвеним мрежама(видети Сliku 3)?



Слика 3: Илустрација аутора лажних вести (Извор: Science, 2018)

4. Потрага за решењем проблема лажних вести

Проблем великих размера привлачи пажњу разних људи и организација који траже решења или макар начине да ублаже штетно дејство проблема. Такав је случај и са лажним вестима. Постоје два основна приступа решавању проблема лажних вести на друштвеним мрежама, а то су: људска интервенција и коришћење алгоритама (Figueira & Oliveira, 2017).

Људском интервенцијом треба да се утврди веродостојност информација које се деле, а овде главну улогу играју такозвани „фект-чекери“, који би у преводу били људи који проверавају чињенице. Неке од познатих светских фектчекерских организација јесу: „Politifact“ , „FactCheck.org“ , „FlackCheck“, „Snopes“ и др. (Berkeley Library, 2022).

С обзиром да алгоритми имају улогу у ширењу лажних вести, претпоставка је да алгоритми треба и да се користе у спречавању тог ширења. Односно коришћење алгоритама у борби са другим (у овом случају штетним) алгоритмима. Алгоритми треба да идентификују лажни садржај и да провере изворе информација. У пракси је упркос разним покушајима за сада тешко створити алгоритме који би били у стању да поуздано утврде да ли је нека информација истинита или лажна (Figueira & Oliveira, 2017) .

Како је у раду већ назначено, једна од већих препрека у спречавању ширења лажних вести јесте огромна количина података коју је практично немогуће „претрести“ без коришћења алатки специјално за тај задатак конструисаних. Горе наведене фектчекерске организације се већ дуже од једне деценије баве овим великим задатком превенције ширења лажних вести. Они врше надзор над друштвеним мрежама, траже потенцијално лажне вести да би их потом или разоткрили или потврдили. Међутим овај посао је јако захтеван и представља

велики интелектуални напор. Постоје процене да је у просеку потребно чак 13 сати да се објави и подели истинита вест након што лажна вест достигне свој врхунац, те она добија и мање пажње од првобитно објављеног дезинформације (Azevedo, 2020). Такође, фектчекери морају бити адекватно тренирани за овај посао, будући да је једно истраживање показало како људи без тренинга у просеку погађају истинитост неке тврдње у тек 54% ситуација (Bond & DePaulo, 2006). Дакле, људски приступ се испоставља као скуп и недовољно ефикасан у борби против ширења лажних вести.

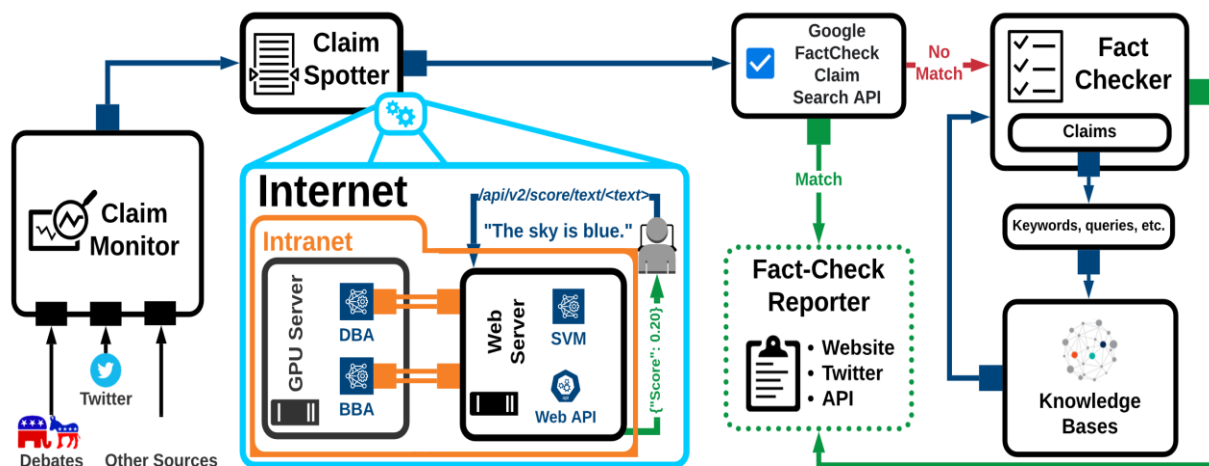
Када је реч о алгоритамском приступу, постоје стандардни начини на који се може приступити откривању лажних вести. Огромна количина података, на супрот када је реч о људском приступу, овде има користи, јер су подаци неопходни за машинско учење. Проблем за алгоритме је сама суштина тврдњи и како програм може разазнати шта је истина, а шта лаж. Због ове суштинске неспособности, алгоритамски приступ планира се из више различитих углова, у покушају да се што ближе „примакне“ решењу. Упрошћено, алгоритми могу да се користе пре свега за вршење надзора над друштвеним мрежама и другим медијима, потом да открију и уоче тврдње које се деле, да тврдње провере тако што их упореде са другим познатим тврдњама и могу да се користе да сами генеришу и деле садржај како би потиснули ефекат ширења неке неистине, ширењем проверених информација (Babakar & Моу, 2016). У даљем тексту биће представљен један од првих система који је створен са намером да обухвати све ове приступе и омогући целовито решење за борбу са дезинформацијама – ClaimBuster.

4.1. *ClaimBuster – Систем за откривање и оспоравање обмањујућих чињеница*

Од децембра 2014. године, група стручњака из области информатике, психологије, компјутерских и комуникационих наука радила је на задатку да створи један целовит систем за откривање и спречавање ширења дезинформација. Свом моделу вештачке интелигенције дали су име ClaimBuster – „онај који хапси тврдње“. Систем користи машинско учење, природну обраду језика и технике испитивања база података како би помогао у разоткривању лажних информација које се деле на друштвеним мрежама, али и на телевизији и на вебсајтовима. Циљ пројекта је да ClaimBuster брзо и ефикасно уочи тврдњу коју вреди проверити, да јој према томе да одређену оцену, потом да је на одређен начин провери и да корисницима пружи истиниту верзију онога што је предмет наведене тврдње. Пројекат је задобио већу пажњу током председничких избора у САД 2016. године, када је и први пут јавно тестиран.

4.1.1. Преглед система ClaimBuster

Систем се састоји од неколико компоненти, које уједно представљају фазе у процесу проверавања тврдњи. Кључне фазе су: мониторинг, уочавање тврдњи, упоређивање тврдњи, проверавање тврдњи и фект-чекерски извештај (видети Слика 4).



Слика 4: Шематски приказ система ClaimBuster (Извор: ClaimBuster, 2022)

Мониторинг – у склопу ове компоненте врши се непрекидан надзор и прикупљање текстуалних садржаја из разних извора, као што су телевизијски преноси, вебсајтови и друштвене мреже. На друштвеној мрежи Твитер ClaimBuster прати хиљаде налога, махом политичких актера у САД.

Уочавање тврдњи – компонента система која се до сада показала најзначајнијом и најкориснијом. Она реченицама додељује оцену од 0.0 до 1.0 у зависности од тога колико је она „чињенична“. Наиме, субјективне и нерелевантне чињеничне реченице добијају нижу оцену, а оне које садрже објективну тврдњу добијају оцену ближу 1.0. Дакле, што је оцена ближа 1.0, то је значајније такву реченицу или боље речено, тврдњу, проверити. Да би ова компонента била успешна, коришћене су технике машинског учења и природне обраде језика. Задатку је приступљено као задатку категоризације и рангирања. Подаци на којима се систем учио јесу све доступне реченице на претходним америчким председничким дебатама, које су стручњаци сврставали у три различите категорије. Те три категорије јесу:

- Субјективне реченице, које не садрже никакву (тачну или нетачну чињеницу), на пример: „Мислим да је тренутак да разговарамо о будућности“;
- Неважне објективне тврдње, то су тврдње које представљају чињенице, али оне које није вредно проверавати јер јавност не интересује да ли су оне тачне или нетачне, на пример: „Пре два дана ручали смо у ресторану“;
- Чињеничне тврдње које вреди проверити, које садрже чињеницу за коју је јавност заинтересована да сазна да ли је тачна или нетачна. Новинари и фект-чекери траже овакве тврдње да би их проверавали. На пример: „Један и по милион америчких грађана болује од ХИВ-а“ (Hassan et al., 2017).

Анализе су показале да постоји висок ниво корелације у избору реченица са америчке председничке дебате 2016. године које је овај алгоритам издвојио да су

вредне провере са онима које су људи у фектчекерским организацијама PolitiFact.com и FactCheck.org издвојили и проверили. Ово је значајан резултат, зато што би стручњацима који се баве спречавањем дезинформисања овакав алат био од велике користи у уштеди времена, а нарочито ако се осврнемо на брзину којом се лажне вести прошире између људи (Hassan et al., 2017a).

Након ове фазе уочавања и оцењивања тврдњи, следи провера тих тврдњи. Систем предвиђа два приступа, упоређивање и проверу.

Упоређивање тврдњи – Након што се тврдња уочи она се најпре упоређује са тврдњама из база података фект-чекерских организација. ClaimBuster претражује базе података и упоређује тврдњу како би пронашао да ли је нека слична или идентична тврдња раније била проверена. Уколико јесте, кориснику се доставља резултат те претходне провере. Позната је тенденција да се сличне лажне вести изнова пласирају и да се не би изнова вршиле провере, овај корак служи да се већ постојећи ресурси искористе да се кориснику достави истина.

Провера тврдњи – У овој фази, тврдња која се проверава се помоћу специјализоване алатке претвара у форму питања. Та питања се уносе као инпут у Wolfram Alpha систем за пружање одговора, али и у Гугл претраживач. Истовремено оригинална тврдња се уноси у Гугл претраживач, а ClaimBuster сачува најрелевантније понуђене странице. На тим страницама, уочава се тврдња, али и њене околне реченице и на тај начин се формира контекст.

Фект-чекерски извештај – Након провере, кориснику се достављају резултати претрага. У три издвојена прозора корисник добија информацију о томе да ли проверавана тврдња постоји у базама података фект-чекерских организација, затим одговоре које пружају Wolfram и Google и контекст пронађен путем Гугл претраживача (Hassan et al., 2017a).

Тешко је проценити и измерити значај овог амбициозног пројекта, посебно ако посматрамо потенцијал који он има да утиче на даље покушаје да се изнађе решење за ширење лажних вести у виду алгоритма. Међутим, ако се посматра искључиво његова тренутна практична и употребна вредност, ClaimBuster није систем који може окончати овај проблем огромних размера. Позитивна страна система је његова ClaimSpotter компонента за уочавање тврдњи, која дефинитивно може убрзати и олакшати рад људима у фект-чекерским организацијама. Целокупан систем и има неку употребну вредност, јер може заинтересованом кориснику помоћи да брзо добије контекст за неку тврдњу, као и да сазна да ли је та тврдња била предмет неких ранијих провера тачности. Негативне стране су у томе што систем не може да ради са фотографским и видео садржајима и потпуно је орјентисан на енглеско говорно подручје, те док се не изврше одређена прилагођавања не може бити користан за проверавање информација на осталим језицима. Додатна слабост је у недостатку прецизности система, али будући да алгоритам не може да разуме било какав контекст и речи као такве не прима са њиховим суштинским значењем, ово је од самог почетка био и остао највећи изазов за алгоритамски приступ борби са лажним вестима.

4.2. Комбиновање људског и алгоритамског приступа у борби са лажним вестима

Наивно је веровати да ће вештачка интелигенција моћи сама да реши проблем лажних вести. Од самог почетка и дефинисања појма лажних вести јасно је да појам захвата многе нијансе – нису све лажне вести у потпуности лажне. Тежак је задатак уопште операционализовати појам лажних вести, а док један програм не познаје у потпуности то шта треба да детектује, неће моћи да без грешке обавља своју функцију. Лажне вести нису математичко питање алгоритама и података (Cassauwers, 2019). Вреди споменути и проблем цензуре и

укидања слободе говора. Чак и људи имају проблем са прецизним погађањем истинитости неких вести, да ли је онда заиста паметно препустити моћ једној вештачкој интелигенцији да означава вести као нетачне и да их цензурише? Овакав систем би у погрешним рукама представљао велики изазов за слободу говора и свакако би захтевао неутралну трећу страну која би вршила надзор над њим (Lee & Fung, 2021).

Са друге стране, у раду је већ назначено да људи без неопходних алата нису у стању да зауставе проток лажних вести, нарочито када их алгоритми друштвених мрежа запазе као значајне и почну да их истичу и на тај начин убрзавају њихово ширење. Због тога је можда и једино решење комбиновати људски и алгоритамски приступ у борби са лажним вестима. Пример људско-технолошког приступа је пројекат спроведен од стране литванског Министарства одбране 2018. године. Направили су програм вештачке интелигенције који је у стању да детектује (потенцијално) лажну вест у року од 2 минута након њеног објављивања, а који потом вест доставља запосленим стручњацима који врше проверу и евалуацију вести (Somer, 2018).

Делује да је у пракси најбољи управо тај приступ: да се користе алгоритми за детекцију, а стручњаци за проверу вести (Cohen, 2020). Такође, веома корисна у спречавању масовног ширења лажних вести била би едукација људи – корисника друштвених мрежа, о конзумирању медија и тумачењу вести; уз то, промоција критичког начина размишљања и подстицање корисника да све информације провере пре него што их поделе са својим пријатељима и пратиоцима.

5. Закључак

Ширење лажних вести је значајан и веома актуелан проблем, чије последице није могуће прецизно утврдити. Оно што је сигурна последица јесте обмањивање људи, стварање илузије, развијање погрешних идеја. Све то може пореметити човека, одвести његово размишљање а потом и живот у неком погрешном правцу. У данашње доба друштвених мрежа повезивање људи са свих страна света уз свега „пар кликова“ избрисало је стварну физичку дистанцу као фактор ограничавања међуљудске интеракције. На интернету су лако доступни речници и веома је лако нечију текстуалну поруку провести кроз „преводац“ и разумети је на матерњем језику без обзира да ли је оригинал био писан на енглеском, руском, кинеском или било ком језику. Другим речима повезивање људи и њихова комуникација данас скоро па нема ограничења. Људи су друштвена бића, морају да припадају групама. У горе наведеним условима долази до повезивања највише на основу занимања, интересовања, ставова, укуса и уверења. Овај процес потпомогнут је механизмима друштвених мрежа које још више олакшавају процес упознавања и на суптилне начине упућују истомишљенике једне на друге. Затим, настају разне групе које се могу повезивати на безброј различитих основа, те се лако могу упознавати и организовати љубитељи филма, неког музичара или фудбалског клуба. Могу се груписати људи на бази сличног смисла за хумор, или политичких уверења, религије и тако даље. Проблем је што је основ груписања најчешће нерегулисан и може бити практично било шта што је људима занимљиво. Системи друштвеног повезивања на интернету су неселективни – они су базирани на безличним алгоритмима и свакако истомишљенике упућују једне на друге независно од природе њихових уверења и ставова. На интернету су карактеристичне појаве разних супкултура, од којих неке могу бити веома корисне или веома штетне по своје чланове или људе изван. Штетне групе најчешће се образују на бази погрешних уверења.

Велику улогу у овом кораку имају дезинформације и лажне вести. На основу лажних информација могуће је изманипулисати одређен део популације који би потом запао у групу истомишљеника где би додатно учврстио своје ставове базиране на лажима. У неким екстремним случајевима ово би могло довести до стварања, примера ради, неких секти или екстремистичких група људи. Са друге стране, мониторинг ових група и пласирање одређених излажираних садржаја члановима групе потенцијално оставља пуно простора за контролу, управљање и манипулисање друштвом. Закључак је да цео систем има огроман потенцијал, а представља релативан новитет и није могуће исконтролисати га, макар не у демократским системима, како не би дошло до укидања слободе говора, као једног од основних загарантованих људских права и слобода. Неопходно је обучити људе кориснике друштвених мрежа, упутити их у проблем лажних вести, освестити их о начину на који друштвене мреже функционишу, како би се осигурало да људи користе друштвене мреже као алат, уместо да сами буду предмет интересних група које препознају потенцијал система или чак предмет безличних алгоритама. Сем тога корисно је подстицати људе на критичко размишљање и проверу вести које читају укључујући и проверу извора вести, пре него што вест поделе. Уз обучене савесне тимове, опремљене потребним алгоритамским системима и одређеним одговорним надзором над њима, могуће је проблем ширења лажних вести на друштвеним мрежама ублажити.

Литература

- Akram, W. & Kumar, R. (2017). A Study on Positive and Negative Effects of Social Media on Society. *International Journal of Computer Sciences and Engineering*, 5(10), 347-354.
- Azevedo, L. (2020). How to beat fake news with algorithms. RTE. Preuzeto 21. aprila 2022. sa : <https://www.rte.ie/brainstorm/2019/0125/1025444-how-to-beat-fake-news-with-algorithms/>
- Babakar, M. & Moy, W., (2016). The State of Automated Factchecking. Full fact. Preuzeto 21. aprila 2022. sa : https://fullfact.org/media/uploads/full_fact-the_state_of_automated_factchecking_aug_2016.pdf
- Berkeley Library. (2022). Real News/ Fake news: Fact Checkers. University of California. Preuzeto 15. aprila 2022. sa: <https://guides.lib.berkeley.edu/c.php?g=620677&p=4333407>
- Bond, C. & DePaulo, B. (2006). Accuracy of deception judgments. *Personality and Social Psychology review: an official journal of the Society for Personality and Social Psychology, Inc.*, 10(3), 214-234.
- Bovet, A. & Makse, H. (2019). Influence of fake news in Twitter during the 2016 presidential election. *Nature Communications*, 10(7), 112-135.
- Carrion-Alvarez, D. & Tijerina-Salina, P. (2020). Fake news in COVID-19: A perspective. *Health Promot Perspect*, 10(4), 290-291.
- Cassauwers, T. (2019). Can artificial intelligence help end fake news?. *Horizon - the EU Research & Innovation Magazine*. Preuzeto 21. aprila 2022. sa: <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news>
- Cohen, I. (2020). Can AI Analytics Weed Out Fake News?. *Anodot*. Preuzeto 1. maja 2022. sa: <https://www.anodot.com/blog/ai-analytics-stops-fake-news/>
- Conover, M., Goncalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alingment of Twitter Users. 2011 IEEE Third Int'l Conference On Privacy, Security, Risk and Trust and 2011 IEEE Thir Int'l Conference On Social Computing.
- Dalomba, F. (2020). Pros and Cons of Social Media. *Lifespan*. Preuzeto 10. aprila, 2021., sa: <https://www.lifespan.org/lifespan-living/social-media-good-bad-and-ugly>
- Dixon, S. (2022). Social Media – Statistics and facts. *Statista*. Preuzeto 7. aprila 2022., sa: <https://www.statista.com/topics/1164/social-networks/>
- Dizikes, P. (2018). Study: On Twitter, false news travels faster than true stories. *MIT News*.

- Preuzeto 15. aprila 2022., sa: <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>
- Duffy, C. & Metz, R. (2022) Why Ukraine War Misinformation is so Hard to Police. CNN Business. Preuzeto 10. aprila 2022. sa: <https://edition.cnn.com/2022/03/15/tech/ukraine-russia-misinformation-challenges/index.html>
- Figueira, A., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121, 817-825.
- Hassan, N., Arslan, F., Li, C. & Tremayne, M. (2017). Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. Preuzeto 15. aprila 2022. sa: <https://ranger.uta.edu/~cli/pubs/2017/claimbuster-kdd17-hassan.pdf>
- Hassan, N., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S., Hasan, S., Joseph, M., Kulkarni, A., Nayak, A., Sable, V., Li, C. & Tremayne, M. (2017). ClaimBuster: The First-ever End-to-end Fact-checking System. Preuzeto 15. aprila sa: <https://ranger.uta.edu/~cli/pubs/2017/claimbuster-vldb17demo-hassan.pdf>
- Howell, L. (2013). Digital wildfires in a hyperconnected world. WEF Report. 3, 15-94. Preuzeto 12. aprila 2022., sa: <https://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/>
- Hughes, H. & Waismel-Manor, I. (2021). The Macedonian Fake News Industry and the 2016 US Election. *PS: Political Science & Politics*, 54(1), 19-23.
- Kent, K. S. (2013). „Propaganda, Public Opinion, and the Second South African Boer War.“ *Inquiries Journal/ Student Pulse* 5(10). Preuzeto 20. decembra, 2021., sa <http://www.inquiriesjournal.com/a?id=781>
- Kreps, S., McCain, R., & Brundage, M. (2022). All the News That's Fit to Fabricate: AI-Generated text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9(1), 104-117.
- Lee, S. & Fung, B. (2021) Artificial intelligence may not actually be the solution for stopping the spread of fake news. *The Conversation*. Preuzeto 1. maja 2022. sa: <https://theconversation.com/artificial-intelligence-may-not-actually-be-the-solution-for-stopping-the-spread-of-fake-news-172001>
- McFadden, C. (2020). A chronological History of Social Media. Interesting engineering. Preuzeto 10. aprila 2021., sa: <https://interestingengineering.com/a-chronological-history-of-social-media>
- Menczer, F. & Hills, T. (2020). Information Overload Helps Fake News Spread, and Social Media Knows It. *The Scientific American* (online) Preuzeto 15. aprila 2022., sa: <https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/>

- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). „Fake news“ Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, 65(2), 180-212.
- Oxford Languages, (2016). Word of the Year 2016. Преузето 15. фебруара 2022., са: <https://languages.oup.com/word-of-the-year/2016/>
- Posetti, J. & Matthews, A. (2018). A short guide to the history of “fake news” and disinformation: A learning module for journalists and journalism educators. ICFJ Internation Center for Journalists. Преузето 21. децембра 2021., са: https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf
- Sawant, P. (2021). Detection & Classification of AI Generated Fake News. *Analytics India Magazine*. Преузето 10. маја 2022. са: <https://analyticsindiamag.com/detection-classification-of-ai-generated-fake-news/>
- Somer, I. (2018). Lithuanians create artificial intelligence with ability to identify fake news in 2 minutes. *Kyiv Post*. Преузето 3. маја 2022. са: <https://www.kyivpost.com/technology/lithuanian-creates-artificial-intelligence-with-ability-to-identify-fake-news-within-2-minutes.html>
- Tandoc, E., Lim, Z., & Ling, R. (2017). Defining “Fake News”. *Digital Journalism*, 6(2), 137-153.
- Thornton, B. (2000). The Moon Hoax: Debates About Ethics in 1835 New York Newspapers. *Journal Of Mass Media Ethics*, 15(2), 89-100.
- Tufekci, Z. (2018). It’s the (Democracy-Poisoning) Golden Age of Free Speech. *Wired*. Преузето 20. децембра, 2021., са : <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/?CNDID=50121752>
- Wang, Q., Chen, W., and Liang, Y. (2011). The Effects of Social Media on College Students. *Johnson & Wales University, Providence, RI*.
- Груловић, Б. (2021). Лажне вести и стварање панике у јавности. *Актуелности*, 39 (?), 57-68.
- Јовановић, Н. и Богдановић, М. (2018). Од гласине до алгоритма. Семинарски рад. Београд: Рачунарство у друштвеним наукама
- Крстић, С. (2018). Шта су друштвене мреже? АТЕС Technologies. Преузето 7. априла 2022., са: <https://atec.rs/sta-su-drustvene-mreze/>
- Ташковић, А. (2022) Леш који се помера и мистериозни пилот из Кијева: Лажне вести о Украјини. *NovaS*. Преузето 10. априла 2022. са: <https://nova.rs/svet/les-koji-se-pomera-i-misteriozni-pilot-iz-kijeva-lazne-vesti-o-ukrajini-foto/>

ИЗЈАВА О АКАДЕМСКОЈ ЧЕСТИТОСТИ

Изјављујем да сам у приложеном раду поштовао/ла сва правила о академској честитости.

Овај писани рад резултат је искључиво мог личног рада, темељи се на мојим истражиањима и ослања се на наведену литературу.

У Београду, дана _____ године.

Потпис студента:
