

# МОГУЋА РЕШЕЊА ЗА САЈБЕР НАСИЉЕ

Татјана ПЕТРОВИЋ, демонстратор

---

\* Факултет безбедности, Универзитет у Београду, petrovictatjana24@gmail.com



## МОГУЋА РЕШЕЊА ЗА САЈБЕР НАСИЉЕ

**Сажетак:** У раду су приказана потенцијална решења за сајбер насиље, односно за различите облике насиља који се испољавају у сајбер простору. Рад превасходно отпочиње пружањем осврта на теоријско одређење, кључне одлике сајбер насиља и улоге које особе могу имати у истом. У том контексту, указано је на специфичности сајбер насиља у односу на традиционално насиље, узимајући у обзир карактеристике самог сајбер простора. Посебна пажња је посвећена разматрању штетних последица сајбер насиља, на основу којих је указано на знатну опасност ове појаве и неопходност њеног адекватног сузбијања и спречавања. С тим у вези, у раду ће бити размотрена могућа решења за сајбер насиље која се односе на примену савремених технологија у сајбер простору. Конкретно, биће истакнут значај информационих технологија у борби против сајбер насиља, кроз разматрање њихове примене у детекцији сајбер насиља на основу анализе текстуалног и другог садржаја. Способности вештачке интелигенције и са њом повезаних области, односно технологија које се примењују у сузбијању сајбер насиља, биће илустриране кроз конкретне пројекте и моделе који су развијани у ове сврхе, а који се примењују и у случајевима вршњачког сајбер насиља. Такође, кроз рад ће бити приказани и могући начини употребе високоразвијених технологија за пружање подршке жртвама сајбер насиља.

**Кључне речи:** сајбер насиље, вршњачко сајбер насиље, вештачка интелигенција, машинско учење, обрада природног језика.

### Увод

Убрзан развој технологија и интернета последњих деценија са собом уједно повлачи и значајну могућност њихове злоупотребе. Управо је сајбер насиље једно од облика штетног понашања која се везују за сајбер простор.

Сајбер насиље је комплексна појава, која се манифестује на различитим платформама и друштвеним мрежама, односно широко је рас-

прострањена у сајбер простору, при чему конкретне манифестације сајбер насиља попримају различите форме или појавне облике. Услед специфичности сајбер простора, сајбер насиље може потенцијално бити штетније од традиционалног насиља. Такође, сајбер насиље је специфично и по питању улога које особе имају у истом, где се поред поменутих сајбер насилника и жртава, јавља категорија посматрача, као веома значајна за вршење, односно сузбијање сајбер насиља.

У раду ће посебна пажња бити посвећена штетности сајбер насиља, односно негативним последицама које сајбер насиље изазива код жртава. Кроз разматрање штетних последица, биће указано на неопходност сузбијања сајбер насиља, при чему ће у раду акценат бити на потенцијалним решењима сајбер насиља која се примењују у сајбер простору. У том контексту, биће приказана примена информационих технологија као потенцијално решење за сајбер насиље. Поред примене у детекцији, али и превенцији сајбер насиља, уједно ће бити указано и на употребу различитих технологија у сврху пружања подршке жртвама сајбер насиља. У раду ће бити размотрена употреба појединих способности вештачке интелигенције и повезаних области у борби против сајбер насиља, при чему ће превасходно бити разматрана примена машинског учења на друштвеним мрежама. Такође, у раду су представљени и комбиновани модели, који су базирани на комбиновању различитих технологија, односно способности вештачке интелигенције са намером постизања веће ефикасност у борби против сајбер насиља. Такође, у раду ће бити приказани и обрада природног језика, софтвери за родитељски надзор, као и конкретни пројекти за сузбијање сајбер насиља који су имплементирани у образовним институцијама. Поред наведених, биће приказана и поједина решења намењена превенцији сајбер насиља, а која се базирају на слању рефлексивних порука.

## Сајбер насиље

Када је у питању дефинисање сајбер насиља, неопходно је указати на чињеницу да тренутно не постоји универзално прихваћена дефиниција ове појаве, те да се сајбер насиље (Cyberbullying) најчешће одређује у складу са Олвеусовом (Olweus) дефиницијом, односно као „насиље које се дешава у сајбер простору употребом компјутера, паметних телефона и других електронских уређаја” (Vismara et al., 2022,

р. 3). Додатно, не постоји ни консензус у погледу адекватног термина за означавање овог вида понашања у сајбер простору, при чему се нпр. у српском језику за сајбер насиље симултано користе термини сајбер булинг, електронско насиље, дигитално насиље и онлајн насиље (Milošević i Putnik, 2019).

Сајбер насиље се манифестује, односно спроводи у сајбер простору, при чему оно није везано за једну конкретну платформу или друштвену мрежу. Тачније, насиље у сајбер простору се испољава преко различитих платформи на интернету, од чет-соба, друштвених мрежа и медија па све до компјутерских игрица (Milošević i Putnik, 2019). Конкретни појавни облици сајбер насиља на овим платформама обухватају широк спектар напада, односно штетног понашања – од вређања, узнемиравања и оговарања, све до лажног представљања, недозвољеног саопштавања, као и обмањивања, искључивања, али и прогањања у сајбер простору (Zlatković i Denić, 2019).

Иако су поједине карактеристике сајбер насиља и традиционалног насиља истоветне, специфичности везане за сам сајбер простор утичу на дистинкцију између традиционалног и сајбер насиља. На пример, према Милошевићу и Путнику, за разлику од традиционалног насиља напад у сајбер простору првенствено омогућује нападачу да остане анониман. Дакле, сајбер насилници су у стању да у потпуности прикрију свој идентитет, употребом лажних имена, података и профила, док се жртва, са друге стране, не може сакрити од сајбер насилника. Додатно, размере потенцијалне публике када је у питању сајбер насиље су несагледиве. Такође, постоји директна повезаност између сајбер и традиционалног насиља у случајевима када сајбер насиље заправо представља својеврсну континуацију традиционалног насиља, сада у сајбер простору (2019). Сајбер насиље је специфично и услед самог начина на који се успоставља и одвија комуникација између насилника и жртве. Наиме, сајбер насиље је потенцијално опасније од традиционалног насиља управо због удаљености сајбер насилника од жртве (Dinić, 2022).

Ради свеобухватнијег разумевања феномена сајбер насиља, неопходно је такође размотрити и улоге које особе могу имати у истом, а које се односе на сајбер насилнике, жртве и посматраче. Пре свега, када су у питању сајбер насилници, Ачутан (Achuthan) и сарадници појашњавају да су код ових особа – нарочито деце која спроводе вршњачко сајбер насиље – пре свега присутне психопатске и макијавелистичке тенденције које се повезују са потребом за доминирањем

или показивањем моћи у групи. Додатно, изражено је и присуство агресије, која се испољава у акцијама сајбер насиља које су репетитивне. Притом, емоционална интелигенција и емпатија су недовољно присутни код сајбер насилника, при чему се развијањем управо ових особина значајно доприноси смањењу насилничког понашања у сајбер простору, али и отпорности жртава (2023). Разлози због којих особа врши сајбер насиље се у великој мери повезују са разлозима вршења традиционалног насиља, дакле најистакнутији мотиви су истоветни и односе се на манифестовање надмоћи насилника над жртвом, стицање одређене позиције у групи, као и одређена негативна осећања и перцепције у вези жртве који су присутни код насилника (Dinić, 2022).

Када су у питању жртве сајбер насиља, у контексту вршњачког сајбер насиља, деца из различитих разлога могу бити изабрана од стране сајбер насилника. Између осталог, сајбер насилници су усмерени и на вршњаке који су се тек доселили, на децу која су стидљива и повучена, богатија или сиромашнија од насилника и сл. (Zlatković i Denić, 2019). Поред наведених, жртве су неретко и популарнија деца и деца која имају шири круг пријатеља, а нарочито деца за коју сајбер насилници сматрају да су слабија од њих. Нарочито се истиче виктимизација деце која су у било ком виду другачија од својих вршњака, дакле може бити речи о деци друге националности, вере или деци која имају физичке или психичке поремећаје (Zlatković i Denić, 2019). Такође, ризик од виктимизације у контексту сајбер насиља је у директној корелацији са количином времена које дете проводи на мобилном телефону или интернету (Kuzmanović i sar., 2016).

Поред поменутог вршњачког сајбер насиља, односно сајбер насиља међу децом, треба напоменути да се виктимизација у контексту сајбер насиља односи и на одрасле особе, како током универзитетског школовања, тако и на радном месту (Achuthan et al., 2023).

У погледу виктимизације, али и вршења сајбер насиља, значајно је указати и на супротстављене ставове по питању утицаја пандемије COVID-19 на поменуте трендове. Наиме, пандемија COVID-19 је између осталог била праћена и својеврсном инфодемијом и свеукупно повећаним ослањањем на интернет и друштвене мреже. Једна група аутора се слаже да су ове околности, поред ризика од стварања зависности од интернета, односно друштвених мрежа, допринеле и интензивирању сајбер насиља током трајања пандемије (Achuthan et al., 2023). Са друге стране, оваквом ставу су супротстављена мишљења

према којим је пандемија COVID-19 потенцијално позитивно утицала на смањење сајбер насиља, услед схватања да су људи проводили више времена са члановима породице и да су на тај начин уједно имали више подршке уколико су били жртве сајбер насиља (Huang et al., 2024). Наравно, ризик од виктимизације у контексту сајбер насиља је већи код особа које су током пандемије проводиле више времена на интернету, односно на друштвеним мрежама (Huang et al., 2024).

Када су у питању улоге у сајбер насиљу, последња улога се односи на посматраче (bystanders). Наиме, посматрачи могу бити значајни у контексту супротстављања сајбер насиљу, кроз учешће у превенцији и интервенцији. У суштини, посматрачи могу значајно допринети активностима које се односе на сузбијање сајбер насиља, и том приликом они заправо напуштају улогу посматрача и почињу активно да учествују у борби против сајбер насиља. Позитиван утицај посматрача у сајбер насиљу се односи пре свега на могућност да заштите жртву и реагују на уочено сајбер насиље (Abaido, 2019). Са друге стране, особе које посматрају сајбер насиље и које се притом укључе у сајбер насиље тако што нпр. проследе одређени садржај који је креиран са намером изазивања штетних последица код мете или жртве сајбер насиља, заправо прелазе у улогу насилника (Dinić, 2022). Посматрачи такође могу остати у пасивној улози, односно неретко се неће укључити у сајбер насиље нити у његово сузбијање (Dinić, 2022).

Сајбер насиље може оставити изразито штетне последице по жртве – како код деце, тако и код одраслих особа. Према Јуу (Ju), штетне последице сајбер насиља се најчешће везују за ментално здравље – од изазивања анксиозности и депресије, до суицидалних мисли. У односу на жртве традиционалног насиља, жртве сајбер насиље су потенцијално изложене већем ризику од самоубиства, уколико се узму у обзир разматране карактеристике сајбер насиља, као и чињеница да сајбер насиље није ограничено временом и простором. Поред штетних последица по ментално здравље, сајбер насиље се такође може негативно одразити и на физичко здравље жртве, при чему су најзаступљенији стомачни проблеми и поремећаји сна. Код деце и младих, односно жртава вршњачког сајбер насиља, уочена је и појава поремећаја у исхрани (2023). Сајбер насиље такође доводи и до појаве психосоцијалних проблема код деце, у виду потешкоћа у социјалној интеракцији са вршњацима и ниског степена самопоуздања (Zlatković i Denić, 2019). Додатно, закључено је да су деца која су била жртве сајбер склонија различитим облицима негативног понашања ван сајбер

простора, као што су бежање од куће, или ношење оружја ради сопствене заштите (Chua et al, 2019). Такође, према Ачутану и сарадницима, сајбер насиље значајно отежава и остваривање циљева одрживог развоја услед његових штетних последица по здравље, економију, родну равноправност, мир и правду, при чему су у погледу вршњачког насиља нарочито истичу његове последице на 3. циљ – добро здравље, и 4. циљ односно квалитетно образовање (2023).

## **Могућа решења за сајбер насиље**

Када се узму у обзир наведене карактеристике сајбер насиља, улоге особа у истом, као и велики потенцијал овог штетног понашања у сајбер простору за изазивање озбиљних последица код жртава, јасно је да је од великог значаја развити одговарајуће начине за решење, односно сузбијање сајбер насиља. У том контексту, у овом делу рада ће бити представљена употреба информационих технологија као потенцијално решење за сајбер насиље, са акцентом на различите примене вештачке интелигенције, као и на неке од актуелних пројеката и модела који се примењују у сузбијању сајбер насиља.

### ***Машинско учење***

Машинско учење (Machine Learning – ML) је често представљено као једно од потенцијалних решења за сајбер насиље, услед значајног доприноса машинског учења у детектовању сајбер насиља путем идентификовања кључних речи које су везане за овај феномен. Наиме, машинско учење као грану или област вештачке интелигенције карактерише способност да систем учи и надаље се развија на основу претходног знања и искуства. У односу на степен самосталности у учењу, односно функционисању, можемо разликовати неколико видова машинског учења (Muneer & Fati, 2020). На пример, предложена је употреба модела за детектовање сајбер насиља на Твитеру (Twitter) који би се заснивао на употреби надгледаног машинског учења и којим би биле обухваћене две технике за издвајање карактеристика (Feature Extraction Technique). Након анализе резултата који су добијени применом наведеног модела, закључено је да је СМО класификатор (Sequential Minimal Optimization (SMO) classifier) био најпрецизнији од коришћених класификатора, односно био је успе-

шан у готово 70% случајева (Muneer & Fati, 2020). Такође, за класификацију текстуалног садржаја у оквиру надгледаног машинског учења често се примењује Classifier Support Vector Machine (SVM) (Muneer & Fati, 2020).

Поред наведеног, развијани су и модели који би се примењивали и на другим друштвеним мрежама, као што је Инстаграм (Instagram), и који би притом детектовали сајбер насиље не само на основу анализе текста (најчешће у форми коментара корисника), већ и на основу анализе слика. У спроведеним експерименталним употребама овог модела, уочено је да се анализом различитих врста садржаја побољшава тачност класификације, конкретно у овом случају СВМ класификатора (Muneer & Fati, 2020).

Поред великог потенцијалног доприноса у сузбијању сајбер насиља, уједно треба имати на уму да је примена вештачке интелигенције, односно машинског учења и алгоритама за аутоматску детекцију и означавање сајбер насиља на друштвеним мрежама суочено са различитим потешкоћама. Машинско учење се нпр. нарочито на Твитеру суочавало са потешкоћама које се односе на анализу садржаја за који је неопходно разумевање ширег контекста и сарказма. Како би се превазишле наведене потешкоће, даљи развој машинског учења је усмерен и на унапређивање алгоритама за надгледано и ненадгледано учење, као и на проширивање база података карактеристика у вези сајбер насиља које се класификују (Achuthan et al., 2023).

### **Комбиновани модели**

Када је у питању идентификовање сајбер насиља, значајно је такође размотрити и модел који предлаже Ковачевић и који се заснива на комбинованој употреби технике вештачке интелигенције која се назива Web mining и визуелизације (2012). Web mining се може дефинисати као „апликација Data mining-а над садржајем, структуром и коришћењем Web ресурса како би се издвојило релевантно знање из података на Web-у”, док је Data mining „итеративни процес тражења нових, скривених информација из велике количине података” (Ковачевић, 2012, стр. 229). Са друге стране, визуелизација се у предложеном моделу комплементарно употребљава са Web mining-ом, како би се приказали резултати процеса Data mining-а спроведеног над подацима који се налазе на Web-у (Kovačević, 2012).

Суштина коришћења Web mining-а и визуелизације у контексту решавања проблема сајбер насиља односи се на њихов допринос разумевању података и издвајању значајних информација из њих. Дакле, првенствено се употребом Web mining-а издвајају информације које су садржане у изузетно великом скупу података, при чему се осим информација долази и до значајних образаца, односно патерна који су садржани у одређеном скупу података. Затим се коришћењем визуелизације резултати Web mining-а представљају на начин погодан за њихово потпуније разумевање, односно информације и патерни се представљају у визуелној форми. Важно је напоменути да управо визуелизација може имати изузетно велики значај, јер не само што се информације и патерни представљају на пријемчив начин, већ визуелизација доприноси уочавању патерна који можда нису били уочени након Web mining-а. Управо наведене чињенице потврђују велики значај који комбинована употреба Web mining-а и визуелизације може имати у откривању релевантних информација и патерна (Kovačević, 2012).

У случају идентификације сајбер насиља, односно сајбер насилника, у употреби су класификација и кластеризација, као две технике Web mining-а. Применом класификације могуће је разврставање нпр. порука на оне чији се садржај доводи у везу са сајбер насиљем, односно на поруке које не садрже сајбер насиље. Кластеризација, са друге стране, може наћи примену у креирању кластера, односно за кластеризацију особе у односу на њено понашање на друштвеној мрежи, у нпр. насилнике, жртве или посматраче (Kovačević, 2012).

Дакле, у предложеном моделу Web mining се примењује за детектовање насилника, док визуелизација доприноси запажању сајбер насилника (Kovačević, 2012).

Поред претходно разматраног, изузетно значајан „комбиновани” модел за детекцију сајбер насиља развили су и Алмомани (Almomani) и сарадници. Овај модел би се користио за детекцију сајбер насиља које је присутно на друштвеним мрежама, при чему је модел замишљен тако да се комбинује употреба одређених способности дубоког учења (Deep Learning – DL) и машинског учења (Machine Learning – ML) (Almomani et al., 2024). Функционисање наведеног модела се одвија кроз неколико фаза, где на почетку процеса слике са друштвених мрежа пролазе кроз кораке претходне обраде, који подразумевају нпр. промену величине слике. Након тога, модел користи способности учења дубоких конволуционих неуралних мрежа (Deep

Convolutional Neural Networks – CNNs). У ту сврху, користе се претходно обучени модели попут VGG16 и ResNet50 за издвајање карактеристика из садржаја. Ови модели се користе за анализу садржаја, односно у овом случају се анализирају слике са друштвених мрежа које су прошле кроз претходну обраду. Кроз анализу слика, ови модели откривају и издвајају унапред одређене карактеристике. Карактеристике до којих се дошло употребом више CNNs-ова се затим комбинују у јединствено заједничко представљање знања. Скуп идентификованих карактеристика се затим користи као инпут за класификаторе машинског учења (нпр. Support Vector Machines или Logistic Regression). На крају, на основу научених образаца односно карактеристика, модел Алмоманија и сарадника се употребљава за категоризацију слика са друштвених мрежа на оне које се могу или не могу сврстати под сајбер насиље (Almomani et al., 2024). Предвиђено је да постоји потенцијал да се модел даље развија тако да обухвати не само слике, односно податке са слика, већ и да се примењује на садржај који обухвата комбинацију текста, звука и видео снимака са друштвених мрежа (Almomani et al., 2024).

### **Обрада природног језика**

За детекцију сајбер насиља се такође користи обрада природног језика (Natural Language Processing – NLP). Када су у питању конкретне примене обраде природног језика као потенцијалног решења за сајбер насиље, потребно је пре свега истаћи да се обрада природног језика често примењује за функционисање четботова (Chatbots), како би се допринело бољој комуникацији између рачунара и корисника (Baltezarević i sar., 2023). Такође, обрада природног језика је коришћена у видео-игрицама, како би се анализирала комуникација међу играчима која подразумева писани текст или говорни садржај. Циљ употребе обраде природног језика у овом контексту јесте откривање сајбер насиља, односно увредљивог садржаја у комуникацији између играча, при чему у овој комуникацији може учествовати велики број особа. Обрада природног језика се комбинује са употребом машинског учења и за припрему, односно обуку алгоритама користе се подаци који се односе на различите увреде, односно негативне коментаре и сличне садржаје сајбер насиља који су присутни у комуникацији међу играчима видео-игрица (Baltezarević i sar., 2023).

Обрада природног језика је такође развијана у правцу адекватне детекције сајбер насиља у случајевима када се користи неколико језика у једној објави на друштвеној мрежи. Пол (Paul) и сарадници су радили на развоју модела који би се користио за детекцију сајбер насиља током пандемије COVID-19 и који би се примењивао на садржају у коме корисници употребљавају два или више језика, односно прелазе са употребе једног језика на друге (code-switch) у својим објавама. У овом моделу је коришћен Deep Ensemble Learning како би се комбиновале способности, тачније инпути машинског и дубинског учења, чиме је развијен јединствен модел за детекцију сајбер насиља на друштвеним мрежама. Суштина овог модела јесте комбиновање предвиђања различитих мета-класификатора који су разрађени употребом неколико алгоритама, како би се свеукупно допринело бољој робусности класификатора (2022).

У контексту употребе обраде природног језика као једног од могућих решења за појаву сајбер насиља, уочено је да се посебан проблем јавља када је у питању идентификовање или сузбијање сајбер насиља које се манифестује у случају слабије коришћених језика (low-resource languages), односно језика који су у мањој мери заступљени на интернету и у литератури (Mahmud et al., 2023).

### ***Софтвери за родитељски надзор***

Један од приступа фокусираних на правовремено идентификовање и сузбијање вршњачког сајбер насиља подразумева праћење активности детета на интернету, посебно на друштвеним мрежама. Осим родитеља, школе се такође могу укључити у ову активност, како би се свеукупно побољшала идентификација и борба против сајбер насиља. Неки од софтвера који су развијени у ове сврхе јесу Content Watch Net Nanny и Net Nanny Social, као и Socialshield (Milošević i Putnik, 2019).

За праћење активности детета на друштвеним мрежама у употреби је и вештачка интелигенција која је развијена у склопу IBM Watson-a. Ова вештачка интелигенција је превасходно усмерена на идентификовање знакова да је дете жртва сајбер насиља и обавештавање родитеља о уоченим активностима, при чему ова вештачка интелигенција такође може да саветује родитеље у вези могућих начина решења проблема сајбер насиља у конкретној ситуацији (Rakhmatov, 2022).

Како би сузбијање вршњачког сајбер насиља било ефективније, развијају се алгоритми за софтвере за родитељски надзор и софтвере којима се блокира садржај који обухвата неку од кључних речи повезаних са сајбер насиљем, који би алгоритмима омогућили прецизније детектовање, односно обраду текста у којем је употребљен сарказам као и одређене скраћенице, псовке и увреде, при чему би били обухваћени и случајевим у којима су одређене речи намерно погрешно написане (Rakhmatov, 2022).

### ***Рефлексивне поруке***

Примена вештачке интелигенције у превенцији сајбер насиља се такође односи и на употребу „рефлексивних порука”, односно употребу вештачке интелигенције како би се особи послала порука којом се указује да садржај који она намерава да пошаље или објави потенцијално спада у сајбер насиље. Намера овакве употребе вештачке интелигенције јесте навођење особе на преиспитивање да ли заиста жели да постави или објави такав садржај (Milošević i sar., 2022). Једна од апликација која функционише по наведеном принципу јесте BullStop апликација која се примењује и за детекцију вршњачког сајбер насиља. BullStop је примарно обучавања уз помоћ података који су прикупљени са Твитера и апликација је осмишљена тако да се употребом вештачке интелигенције пре свега уоче и блокирају поруке чији садржај може означити као сајбер насиље, при чему апликација такође има развијене могућности за слање „рефлексивних порука”. Такође, замишљено је да BullStop саветује родитеље и особље у образовним установама како да поступају у конкретном случају сајбер насиља (Milošević i sar., 2022).

Rething Words је средство које се користи како би се допринело превенцији сајбер насиља, а које такође функционише по принципу утицања на слање поруке, пре него што се оне пошаљу. Наиме, ова виртуелна тастатура првенствено онемогућава слање поруке у чијем садржају је детектовано сајбер насиље. Особи се затим упућује обавештење да је у питању увредљив садржај и да особа треба да измени поруку. Rething Words неће онемогућити слање поруке, иако особа не измени њен садржај. Међутим, важно је имати на уму да је према подацима истраживања у коме је анализирана употреба наведене виртуелне тастатуре закључено да је у 93% случајева особа изменила садржај поруке након што јој је скренута пажња на потенцијално сајбер насиље (Kovačević i Bajramović, 2018).

### ***CREEP пројекат***

У контексту употребе информационих технологија у сузбијању сајбер насиља, а нарочито вршњачког сајбер насиља, значајан је и CREEP пројекат. CREEP пројекат је намењен превенцији вршњачког сајбер насиља и заправо обухвата две врсте технологија – CREEP семантичку технологију (CREEP Semantic Technology) и CREEP виртуелног тренера (CREEP Virtual Coach). CREEP семантичка технологија (CREEP Semantic Technology) је платформа која је намењена за анализу садржаја на друштвеним мрежама. Ова платформа је настала како би се омогућила употреба вештачке интелигенције као средства за детектовање употребе говора мржње и сајбер насиља и пре свега је замишљено да се ова платформа имплементира у школама. Замишљено је да се ова платформа користи као средство за правовремено детектовање појаве сајбер насиља међу вршњацима, како би одговорна лица у школама могла да се укључе и предузму одговарајуће мере (Gabrielli et al., 2021).

Као што је наведено, у оквиру CREEP пројекта за сузбијање, односно детектовање сајбер насиља, развијен је и CREEP виртуелни тренер (CREEP Virtual Coach). У овом случају реч је о апликацији за паметне мобилне телефоне која је намењена комуникацији са младима који су изложени сајбер насиљу. Наиме, апликација се користи како би младима пружила подршку да нпр. пријаве сајбер насиље и да се лакше носе са његовим штетним последицама. CREEP виртуелни тренер функционише тако што се млади уз помоћ упитника едукују да препознају различите облике сајбер насиља којима су били изложени као жртве или сведоци, односно посматрачи. Након што се идентификује да ли је у питању сајбер насиље, и о ком облику испољавања сајбер насиља је реч, CREEP виртуелни тренер затим саветује младе како да поступају у конкретной ситуацији. Такође, ова технологија пружа подршку младима који су били жртве сајбер насиља кроз прослеђивање различитих едукативних материјала о начинима како се носити са сајбер насиљем и како појачати сопствену отпорност. Апликација се такође примењује у и превенцији, односно пре него што дође до виктимизације младих (Gabrielli et al., 2021).

## Закључак

Сајбер насиље је вид штетног понашања којим се злоупотребава развој нових технологија и перципиране предности сајбер простора, како би се потенцијално нанела значајна штета, односно последице жртвама сајбер насиља. Стога, неопходно је пре свега поседовати релевантно разумевање самог феномена и његових основних одлика, како би се надаље развијала адекватна решења за појаву сајбер насиља. Како је у раду приказано, тренутно је изостала универзално прихваћена дефиниција сајбер насиља, чиме је не само отежано разумевање сајбер насиља, већ и потенцијална шира сарадња на сузбијању ове веома штетне појаве.

Такође, у раду је указано на специфичности сајбер насиља, односно на одређене карактеристике сајбер простора чија злоупотреба доприноси ставу да је сајбер насиље опасније од традиционалног, услед пре свега дистанце између насилника и жртве, као и чињеници да је сајбер насиље присутно на различитим платформама, и да није ограничено просторним или временским факторима.

Услед велике опасности и тешких последица које сајбер насиље изазива код жртава, јасно је да је неопходно развити правремене и адекватне одговоре на сајбер насиље, те се међу могућим решењима истичу различити модели, пројекти и платформе, односно употреба различитих технологија у сајбер простору.

У раду је посебна пажња посвећена технологијама попут вештачке интелигенције, машинског учења, дубоког учења, обради природног језика и конкретним моделима и платформама као могућим решењима за сајбер насиље. Приказана је примена појединих високоразвијених технологија у детекцији сајбер насиља путем текстуалног садржаја, али су представљени и модели који се односе и на друге садржаје, као што су нпр. слике, при чему је појашњена примена машинског учења и различитих модела за детекцију сајбер насиља на различитим друштвеним мрежама.

Такође, радом су обухваћени и комбиновани модели за сузбијање сајбер насиља, при чему би се у првом моделу користили Web mining и визуелизација, ради детектовања и уочавања сајбер насилника, док се у другом моделу комбинују способности машинског и дубоког учења. Овакав приступ у великој мери доприноси превазилажењу појединих недостатака вештачке интелигенције, односно комплементарном употребом различитих технологија се постиже значајан напредак у

сузбијању сајбер насиља. Такође, као потенцијално решење за сајбер насиље се појављује и употреба обраде природног језика, нарочито када се користи у комбинацији са нпр. машинским учењем.

Када је у питању вршњачко сајбер насиље, у раду је представљено како различити пројекти – попут CREEP пројекта, као и софтвери за родитељски надзор, доприносе сузбијању вршњачког сајбер насиља са једне стране, и подршци жртвама, са друге.

Напошетку, када се узму у обзир карактеристике и штетне последице сајбер насиља, може се закључити да се као једно од могућих решења истиче употреба информационих технологија, односно њена примена у сајбер простору. Како се нове технологије развијају несагледивом брзином, потребно је са једне стране обратити пажњу на могућност њихове злоупотребе, док се такође мора узети у обзир и потенцијал ових технологија за даљи развој одговора на штетна понашања у сајбер простору.

## Литература

- Abaido, G. M. (2019). Cyberbullying on Social Media Platforms Among University Students in the United Arab Emirates. *International Journal of Adolescence and Youth*, 25(1), 407–420. Retrieved from: <https://www.tandfonline.com/doi/full/10.1080/02673843.2019.1669059>
- Achuthan, K., Nair, V. K., Kowalski, R., Ramanathan, S., & Raman, R. (2023). Cyberbullying Research – Alignment to Sustainable Development and Impact of COVID-19: Bibliometrics and Science Mapping Analysis. *Computers in Human Behavior*, 140, 1–17.
- Almomani, A., Nahar, K., Alauthman, M., Al-Betar, M. A., Yaseen, Q., & Gupta, B. B. (2024). Image Cyberbullying Detection and Recognition Using Transfer Deep Machine Learning. *International Journal of Cognitive Computing in Engineering*, 5, 14–26.
- Baltezarević, V., Baltezarević, R. i Baltezarević, I. (2023). Sajber uznemiravanje dece sa posebnim osvrtom na digitalne igre. *Temida*, (26)2, 261–284.
- Chua, Y. T., Parkin, S., Edwards, M., Oliveira, D., Schiffner, S., Tyson, G., & Hutchings, A. (2019). Identifying Unintended Harms of Cybersecurity Countermeasures. In: *2019 APWG Symposium on Electronic Crime Research (eCrime)* (p. 1–15). New York: Institute of Electrical and Electronics Engineers.
- Dinić, B. (2022). *Digitalno nasilje*. Novi Sad: Filozofski fakultet u Novom Sadu.
- Gabrielli, S., Rizzi, S., Carbone, S., & Piras, E. M. (2021). School Interventions for Bullying-Cyberbullying Prevention in Adolescents: Insights from the UPRIGHT and CREEP Projects. *International journal of environmental research and public health*, 18(21), 1–13.

- Huang, N., Zhang, S., Mu, Y., Yu, Y., Riem, M. M. E., & Guo, J. (2024). Does the COVID-19 Pandemic Increase or Decrease the Global Cyberbullying Behaviors? A Systematic Review and Meta-Analysis. *Trauma, Violence & Abuse*, 25(2), 1018–1035.
- Ju, B. (2023). Impacts of Cyberbullying and Its Solutions. *Advances in Humanities Research*, 29, 254–258.
- Kovačević, A. (2012). Primena web mininga i vizualizacije u otkrivanju sajber nasilja. U: *Reagovanje na bezbednosne rizike u obrazovno-vaspitnim ustanovama* (225–241). Beograd: Fakultet bezbednosti.
- Kovačević, A., i Bajramović, J. (2018). Monitoring sajber nasilja. U: *Bezbednost u obrazovno-vaspitnim ustanovama i video-nadzor* (str. 111–126). Beograd: Fakultet bezbednosti.
- Kuzmanović, D., Lajović, B., Grujić, S. i Medenica, G. (2016). *Digitalno nasilje – prevencija i reagovanje*. Beograd: Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije i Pedagoško društvo Srbije.
- Mahmud, T., Ptaszynski, M., Eronen, J., & Masui, F. (2023). Cyberbullying Detection for Low-Resource Languages and Dialects: Review of the State of the Art. *Information Processing & Management*, 60(5), 1–52.
- Milošević, M., i Putnik, N. (2019). Fenomen vršnjačkog nasilja na internetu – izazov za nauku i zakonodavstvo. *Sociologija*, 61(4), 599–616.
- Milošević, T., Van Royen, K. & Davis, B. (2022). Artificial Intelligence to Address Cyberbullying, Harassment and Abuse: New Directions in the Midst of Complexity. *International Journal of Bullying Prevention* 4, 1–5.
- Muneer, A., & Fati, S. M. (2020). A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter. *Future Internet*, 12(11), 187.
- Paul, S., Saha, S., & Singh, J. P. (2023). COVID-19 and Cyberbullying: Deep Ensemble Model to identify Cyberbullying from Code-Switched Languages during the Pandemic. *Multimedia Tools and Applications*, 82(6), 8773–8789.
- Rakhmatov, D. (2022). Methods and Effectiveness of the Use of Artificial Intelligence in the Fight Against Cyberbullying. *Journal of Academic Research and Trends in Educational Sciences*, 1(4), 122–129.
- Vismara, M., Girone, N., Conti, D., Nicolini, G., & Dell’Osso, B. (2022). The Current Status of Cyberbullying Research: A Short Review of the Literature. *Current Opinion in Behavioral Sciences*, 46, 1–17.
- Zlatković, D. i Denić, N. (2019). Prevencija elektronskog nasilja. U: *Zbornik radova sa 5. Međunarodnog savetovanja o upravljanju znanjem i informatici* (str. 65–75). Novi Sad: Visoka tehnička škola strukovnih studija u Novom Sadu.

## POSSIBLE SOLUTIONS FOR CYBERBULLYING

Tatjana PETROVIĆ, Teaching Aide

### *Summary*

The paper presents potential solutions for cyber violence, that is, the solutions for different forms of violence that are manifested in the cyber space. The paper begins with an overview of the theoretical framework, key features of cyberbullying and the roles that people can play in it. In this context, the paper points out the specifics of cyberbullying compared to the traditional bullying, while pointing out the relevant characteristics of the cyber space itself. Based on the consideration of the harmful consequences of cyberbullying, the paper points out the considerable danger of this phenomenon, and consequently, the necessity of its adequate suppression and prevention. In this regard, the paper will present possible solutions for cyber violence related to the application of modern technologies in cyber space. In particular, the paper discusses the importance of utilising information technologies in the fight against cyberbullying, through consideration of their application in the detection of cyberbullying based on the analysis of textual and other content. The capabilities of artificial intelligence and its related fields, that is, technologies that are applied in the suppression of cyber violence will be illustrated through specific projects and models developed for these purposes, which are also applied in cases of peer cyber violence. The paper also examines the possible ways of using certain technologies for providing support to victims of cyber violence.

**Keywords:** *cyberbullying, peer cyberbullying, artificial intelligence, machine learning, natural language processing.*