Pregledni naučni rad
UDK:343.9:004.738.5
DOI:10.7251/ZND2601560G

BEZBJEDNOST I CYBERBULLYING

Dr.sc. Mirko Goreta, viši predavač
Veleučilište Marko Marulić u Kninu[298]

Abstrakt: *Kibernetičko nasilje postalo je raširen problem u modernim digitalnim komunikacijskim okruženjima. Platforme društvenih medija, aplikacije za razmjenu poruka i online zajednice omogućuju pojedincima interakciju putem velikih mreža, ali iste te tehnologije mogu omogućiti i uznemiravanje, zastrašivanje i uporno online zlostavljanje. Za razliku od tradicionalnog nasilja, kibernetičko nasilje često je pojačano karakteristikama digitalnih platformi, uključujući anonimnost, brzo dijeljenje informacija i trajnu vidljivost sadržaja. Ovaj rad ispituje kibernetičko nasilje iz perspektive digitalne sigurnosti i dizajna online platformi. Analizira mehanizme putem kojih se kibernetičko nasilje razvija u digitalnim okruženjima, psihološke i sociološke faktore koji utječu na agresivno online ponašanje te tehnološke sustave koji se koriste za otkrivanje štetnih interakcija. Posebna se pozornost posvećuje mehanizmima kibernetičke sigurnosti kao što su automatizirani sustavi moderiranja, alati za praćenje ponašanja i strukture upravljanja platformama koje imaju za cilj smanjenje uznemiravanja u online prostorima. Analiza pokazuje da kibernetičko nasilje proizlazi iz interakcije između ponašanja korisnika i tehnološke arhitekture digitalnih platformi. Rješavanje ovog problema stoga zahtijeva integrirani pristup koji kombinira tehnologije sigurnosti, odgovorno upravljanje platformama i veću svijest o praksama digitalne komunikacije. Jačanje ovih elemenata može doprinijeti sigurnijim online okruženjima gdje se štetno ponašanje učinkovitije otkriva, ograničava i sprječava.*

Ključne riječi: *kibernetičko nasilje, kibernetička sigurnost, digitalne platforme, moderiranje sadržaja, upravljanje platformama.*

Uvod

Brzo širenje digitalnih komunikacijskih tehnologija transformiralo je način na koji pojedinci međusobno komuniciraju, dijele informacije i sudjeluju u društvenom životu. Platforme društvenih mreža, aplikacije za razmjenu poruka, online forumi i okruženja za igrice stvorili su prostore u kojima se komunikacija odvija konstantno i preko geografskih granica. Iako ove tehnologije pružaju mogućnosti za suradnju i povezivanje, uvele su i nove oblike štetnog ponašanja. Jedan od najznačajnijih je cyberbullying, oblik uznemiravanja koji se događa putem digitalnih komunikacijskih kanala i može uključivati ponovljeno neprijateljstvo, zastrašivanje, ponižavanje ili prijetnje upućene pojedincima putem online platformi (Kowalski, Limber i Agatston, 2012). Kibernetičko nasilje je sve značajnija tema zbog svojih psiholoških, socijalnih i tehnoloških dimenzija. Istraživanja u obrazovnoj psihologiji pokazuju da žrtve kibernetičkog nasilja često doživljavaju tjeskobu, depresiju, smanjen akademski uspjeh i socijalno povlačenje. U svom opsežnom pregledu istraživanja elektroničkog nasilja, Kowalski, Giumetti, Schroeder i Lattanner (2014.) analizirali su podatke iz više empirijskih studija koje su uključivale adolescente i mlade odrasle osobe te pronašli dosljedne veze između viktimizacije kibernetičkim nasiljem i negativnih ishoda mentalnog zdravlja, uključujući povećani

[298] mirko.goreta@veleknin.hr

stres i smanjeno zadovoljstvo životom. Istovremeno, elektroničko nasilje usko je povezano s tehnološkim strukturama digitalnih platformi. Online okruženja omogućuju korisnicima brzu i anonimnu komunikaciju. Ove osobine mogu olakšati štetno ponašanje jer smanjuju društvenu odgovornost. Psiholog John Suler uveo je koncept efekta online dezinhibicije kako bi objasnio zašto se pojedinci često ponašaju agresivnije ili neodgovornije online nego što bi se ponašali u interakciji licem u lice. Prema Suleru, anonimnost, nevidljivost i odsutnost neposrednih društvenih posljedica mogu smanjiti samokontrolu i potaknuti neprijateljsku komunikaciju (Suler, 2004). Ovaj fenomen ilustrira kako tehnološki dizajn djeluje u interakciji s ljudskim ponašanjem stvarajući okruženja u kojima nasilje može cvjetati. Budući da se kibernetičko nasilje događa unutar digitalnih sustava, pristupi kibernetičkoj sigurnosti sve se više smatraju dijelom rješenja. Mehanizmi online sigurnosti poput algoritama za moderiranje sadržaja, sustava za prijavljivanje, alata za provjeru identiteta i politika upravljanja platformama osmišljeni su za otkrivanje štetnog ponašanja i smanjenje njegovog širenja. Studije sustava moderiranja platformi pokazuju da automatizirana jezična analiza i praćenje ponašanja mogu identificirati obrasce zlostavljačkog sadržaja i označiti ih za pregled prije nego što eskaliraju u velike kampanje uznemiravanja (Chandrasekharan i sur., 2017). Ovi tehnološki pristupi pokazuju da prevencija kibernetičkog nasilja nije ograničena samo na društvenu intervenciju, već uključuje i dizajn sigurnih i odgovornih digitalnih okruženja. Međutim, sama tehnološka rješenja ne mogu eliminirati kibernetičko nasilje. Štetno online ponašanje također je oblikovano društvenim normama, grupnom dinamikom i kulturnim očekivanjima unutar digitalnih zajednica. Sociološka istraživanja pokazuju da se nasilničko ponašanje često događa unutar vršnjačkih skupina gdje agresija postaje normalizirana ili nagrađena društvenom pažnjom (Patchin & Hinduja, 2015). To znači da učinkovita prevencija zahtijeva kombinaciju mehanizama kibernetičke sigurnosti, upravljanja platformama i obrazovnih strategija koje potiču odgovorno digitalno ponašanje. Ovaj rad stoga ispituje kibernetičko nasilje iz perspektive digitalne sigurnosti i društvenog ponašanja. Istražuje kako se kibernetičko nasilje pojavljuje u online okruženjima, kako strukture platformi i anonimnost utječu na agresivnu komunikaciju te kako alati kibernetičke sigurnosti i mehanizmi upravljanja mogu pomoći u otkrivanju i sprječavanju štetnog ponašanja. Rad ima za cilj pokazati da rješavanje kibernetičkog nasilja zahtijeva integrirani pristup koji uzima u obzir i dizajn digitalnih sigurnosnih sustava i ponašanje pojedinaca koji ih koriste.

Priroda nasilja u digitalnim okruženjima

Kibernetičko nasilje odnosi se na neprijateljsko ili agresivno ponašanje koje se provodi putem digitalnih komunikacijskih tehnologija s namjerom nanošenja štete drugoj osobi. Za razliku od tradicionalnog nasilja, koje se obično događa u fizičkim okruženjima poput škola ili radnih mjesta, kibernetičko nasilje djeluje putem online platformi gdje komunikacija može dosegnuti široku publiku i ostati dostupna dulje vrijeme. Istraživači općenito definiraju kibernetičko nasilje kao ponovljenu i namjernu štetu nanesenu elektroničkim sredstvima, uključujući društvene mreže, usluge razmjene poruka, platforme za igre ili online forume (Kowalski, Limber i Agatston, 2012). Digitalna priroda ovih interakcija znači da se štetne poruke, slike ili videozapisi mogu brzo distribuirati i teško ih je ukloniti, što često pojačava psihološki utjecaj na žrtve. Kibernetičko nasilje poprima nekoliko prepoznatljivih oblika, od kojih svaki zloupotrebljava osobine digitalnih komunikacijskih sustava. Jedan uobičajeni oblik je online uznemiravanje, koje uključuje ponovljene uvredljive poruke, prijetnje ili uvrede usmjerene prema određenoj osobi. Drugi oblik je doxxing, gdje se privatni ili osjetljivi podaci o osobi javno objavljuju na internetu kako bi se ta osoba osramotila ili zastrašila. Lažno predstavljanje predstavlja još jedan mehanizam u kojem agresor dobiva pristup ili stvara profil koji izgleda kao da pripada žrtvi i koristi ga za objavljivanje štetnog ili obmanjujućeg sadržaja. Istraživači koji proučavaju digitalno uznemiravanje pokazali su da se ove prakse često događaju zajedno, a ne

izolirano, stvarajući kontinuirane kampanje zlostavljanja koje se protežu na više platformi (Hinduja & Patchin, 2014). Ključna karakteristika cyberbullyinga je uloga mrežnog pojačanja. Digitalne platforme omogućuju brzo širenje štetnog sadržaja putem mehanizama dijeljenja kao što su ponovno objavljivanje, komentiranje i algoritamski sustavi preporuka. U mnogim slučajevima, izvorni agresor samo je jedan sudionik u većem procesu u kojem drugi korisnici doprinose ponižavanju ili uznemiravanju žrtve. Studije interakcija na društvenim mrežama pokazuju da zlostavljačke poruke često privlače dodatne sudionike koji se pridružuju uznemiravanju nakon što postane vidljivo unutar online zajednica (Chandrasekharan i sur., 2017). Ovaj učinak pojačanja povećava opseg nasilja izvan onoga što se obično događa u izvanmrežnim okruženjima. Drugi mehanizam koji razlikuje virtualno nasilje od tradicionalnog nasilja je trajnost. U fizičkim okruženjima, štetno ponašanje često se događa unutar ograničenih društvenih konteksta i vremenskih okvira. Međutim, online interakcije stvaraju digitalne zapise koji ostaju dostupni dugo nakon izvornog događaja. Fotografije, komentari ili objave mogu se više puta pregledavati i dijeliti, produžujući izloženost žrtava ponižavanju ili uznemiravanju. Istraživanje digitalne viktimizacije pokazuje da ova upornost značajno povećava psihološki stres jer pojedinci ne mogu lako izbjeći ili kontrolirati širenje štetnog sadržaja (Slonje, Smith i Frisén, 2013). Struktura digitalnih platformi također utječe na razvoj virtualnog nasilja. Osobine poput anonimnih profila, privatnih kanala za razmjenu poruka i mreža velike publike smanjuju izravnu društvenu odgovornost i olakšavaju nasilnicima ciljanje žrtava bez neposrednih posljedica. Te iste osobine mogu zakomplicirati otkrivanje i intervenciju moderatora ili administratora platforme. Kao rezultat toga, virtualno nasilje ne proizlazi samo iz individualnog neprijateljstva već i iz interakcije između ponašanja korisnika i samog dizajna platforme.

Anonimnost, digitalni identitet i psihologija online agresije

Jedan od najčešće raspravljanih faktora koji doprinose kibernetičkom nasilju je struktura digitalnog identiteta u online okruženjima. Mnoge online platforme omogućuju korisnicima interakciju putem pseudonima, anonimnih profila ili profila koji otkrivaju malo osobnih podataka. Iako anonimnost može zaštititi privatnost i potaknuti otvorenu komunikaciju, ona također može smanjiti osobnu odgovornost i učiniti agresivno ponašanje vjerojatnijim. Psiholog John Suler ispitao je ovaj fenomen u svojoj studiji o učinku online dezinhibicije, koji opisuje kako se pojedinci ponašaju slobodnije, a ponekad i agresivnije u online okruženjima nego u interakciji licem u lice. Suler objašnjava da anonimnost, nevidljivost i odsutnost neposredne društvene povratne informacije mogu oslabiti samokontrolu i učiniti korisnike manje svjesnima posljedica svojih postupaka (Suler, 2004). U digitalnoj komunikaciji pojedinci često ne vide emocionalne reakcije onih na koje ciljaju, što smanjuje empatiju i može potaknuti neprijateljsko ponašanje. Struktura online identiteta također utječe na to kako pojedinci doživljavaju odgovornost za svoje postupke. Sociološka istraživanja deindividuacije sugeriraju da kada se ljudi osjećaju manje osobno prepoznatljivima unutar grupe, veća je vjerojatnost da će se upustiti u ponašanje koje krši društvene norme. Ovaj koncept primijenjen je na digitalna okruženja u kojima korisnici komuniciraju putem velikih mreža i mogu se osjećati psihološki udaljenima od onih kojima nanose štetu. Drugi važan faktor je fragmentacija digitalnog identiteta na više platformi. Pojedinci često održavaju različite profile na društvenim mrežama, forumima i platformama za razmjenu poruka. Svako okruženje može poticati različite oblike interakcije i razine anonimnosti. Znanstvenici koji proučavaju digitalnu komunikaciju otkrili su da ti fragmentirani identiteti omogućuju korisnicima eksperimentiranje s ponašanjem koje bi mogli izbjegavati u offline okruženjima, uključujući neprijateljstvo ili uznemiravanje. Turkleova analiza online zajednica ilustrira kako digitalne platforme omogućuju pojedincima da usvoje alternativne identitete i uloge, što ponekad dovodi do komunikacijskih stilova koji se značajno razlikuju od njihovog ponašanja u stvarnom svijetu (Turkle, 2011). U određenim slučajevima, ovo eksperimentiranje može se razviti u održive

obrasce agresivne komunikacije. Grupna dinamika unutar online zajednica može dodatno pojačati agresivno ponašanje. Platforme društvenih medija često nagrađuju provokativan sadržaj povećanom vidljivošću putem lajkova, dijeljenja ili algoritamske promocije. Ova vidljivost može potaknuti pojedince da sudjeluju u uznemiravanju jer takvo ponašanje privlači pažnju i društveno pojačanje od drugih unutar mreže. Istraživanja koja ispituju online odjeljke za komentare i forume za raspravu pokazala su da agresivne objave često izazivaju dodatne neprijateljske reakcije drugih korisnika, stvarajući kaskadu negativnih interakcija (Chandrasekharan i sur., 2017). Ovi obrasci ilustriraju kako okruženja platformi mogu pojačati pojedinačne činove neprijateljstva u kolektivno uznemiravanje. Psihološka distanca koju stvara digitalna komunikacija također igra ulogu u smanjenju empatije. U komunikaciji licem u lice, pojedinci mogu promatrati ton glasa, izraze lica i govor tijela, koji daju znakove o tome kako se drugi osjećaju. Online komunikacija često uklanja te znakove, što pojedincima olakšava previdjeti emocionalne posljedice svojih poruka. Studije komunikacijske psihologije pokazuju da odsutnost ovih društvenih signala može navesti korisnike da apstraktnije tumače interakcije, fokusirajući se na poruku, a ne na osobu koja je prima (Walther, 2011). Ova smanjena emocionalna svijest doprinosi postojanosti neprijateljske komunikacije u digitalnim okruženjima.

Sigurnosni mehanizmi za otkrivanje i sprječavanje kibernetičkog nasilja

Budući da se kibernetičko nasilje događa unutar digitalnih platformi, tehnološki sigurnosni sustavi igraju sve važniju ulogu u identificiranju i ograničavanju štetnog online ponašanja. Mehanizmi sigurnosti osmišljeni za digitalin prostor djeluju praćenjem digitalne komunikacije, otkrivanjem obrazaca povezanih s uznemiravanjem i omogućavanjem brzih odgovora administratora platforme. Ovi sustavi kombiniraju automatizirane tehnologije, strukture moderiranja i alate za izvještavanje kako bi se smanjilo širenje zlostavljačkog sadržaja. Istraživanja u upravljanju online platformama pokazuju da tehnološki sustavi praćenja mogu značajno smanjiti vidljivost štetne komunikacije kada su integrirani s dosljednim politikama provedbe (Gillespie, 2018). Jedan od najčešće korištenih mehanizama za otkrivanje kibernetičkog nasilja je automatizirano moderiranje sadržaja temeljeno na strojnom učenju i obradi prirodnog jezika. Ovi sustavi analiziraju velike količine online teksta kako bi identificirali jezične obrasce povezane s uznemiravanjem, prijetnjama ili govorom mržnje. Algoritmi se obučavaju na skupovima podataka koji sadrže primjere zlostavljačke i nezlostavljačke komunikacije kako bi mogli prepoznati uvredljive izraze, ponovljeno ciljanje pojedinaca ili koordinirane kampanje uznemiravanja. Studija Davidsona, Warmsleyja, Macyja i Webera ispitala je kako modeli strojnog učenja mogu identificirati govor mržnje i uvredljiv jezik na platformama društvenih medija. Analizirajući tisuće Twitter poruka, istraživači su pokazali da automatizirani klasifikatori mogu razlikovati uvredljiv jezik od istinskog govora mržnje s visokom točnošću kada se obučavaju s pažljivo označenim podacima (Davidson, 2017). Ovi sustavi omogućuju platformama da označe štetan sadržaj za pregled prije nego što se široko proširi digitalnim mrežama. Još jedan važan mehanizam kibernetičke sigurnosti uključuje sustave za prijavljivanje korisnika i moderiranje. Većina glavnih digitalnih platformi pruža alate koji korisnicima omogućuju prijavljivanje uvredljivih poruka, prijetećeg ponašanja ili uznemiravanja. Ove prijave pregledavaju ili ljudski moderatori ili automatizirani sustavi koji daju prioritet najtežim slučajevima. Prisutnost alata za prijavljivanje može obeshrabriti štetno ponašanje jer korisnici znaju da uvredljive radnje mogu rezultirati suspenzijom ili uklanjanjem profila. Istraživanje koje ispituje prakse moderiranja na platformama društvenih medija pokazuje da se učinkoviti sustavi za prijavljivanje oslanjaju na kombinaciju automatiziranog filtriranja i ljudske procjene kako bi se osigurala pravednost i kontekstualna prosudba (Gillespie, 2018). Ljudski moderatori ostaju važni jer automatizirani sustavi ne mogu uvijek interpretirati sarkazam, kulturne reference ili složeni konverzacijski kontekst.

Strategije kibernetičke sigurnosti također uključuju sustave sigurnosti računa i upravljanja identitetom osmišljene za smanjenje zlostavljanja temeljenog na anonimnosti. Iako potpuno otkrivanje identiteta nije uvijek prikladno iz razloga privatnosti, platforme sve više primjenjuju mjere koje ograničavaju stvaranje velikog broja anonimnih računa. Ove mjere uključuju provjeru telefonskog broja, praćenje ponašanja sumnjivih računa i sustave ograničavanja brzine slanja poruka koji sprječavaju korisnike da šalju velike količine poruka u kratkim razdobljima. Studije koordiniranog online uznemiravanja pokazuju da ograničavanje automatiziranog stvaranja računa i sumnjivih aktivnosti može značajno smanjiti sposobnost grupa da organiziraju ciljane kampanje zlostavljanja (Jhaver, Ghoshal, Bruckman i Gilbert, 2019). Praćenje na razini mreže predstavlja još jedan tehnološki pristup sprječavanju kibernetičkog nasilja. Sigurnosni sustavi mogu analizirati obrasce komunikacije na mrežama kako bi otkrili neobične poraste neprijateljskih poruka usmjerenih prema određenoj osobi ili grupi. Na primjer, ako korisnik iznenada primi stotine uvredljivih komentara u kratkom razdoblju, automatizirani sustavi praćenja mogu privremeno ograničiti aktivnost objavljivanja ili upozoriti moderatore da istraže situaciju. Istraživanja o otkrivanju online uznemiravanja pokazala su da su takve tehnike analize mreže posebno korisne za identificiranje koordiniranih događaja uznemiravanja koji uključuju više računa usmjerenih na jednu žrtvu (Chandrasekharan, 2017). Osim tehnologija otkrivanja, mjere kibernetičke sigurnosti uključuju i zaštitne alate za korisnike. Mnoge platforme pružaju opcije poput blokiranja ili filtriranja poruka koje pojedincima omogućuju ograničavanje interakcija s nasilnim korisnicima. Iako ovi alati ne uklanjaju u potpunosti kibernetičko nasilje, pomažu žrtvama da ponovno preuzmu kontrolu nad svojim online okruženjem. Studije koje ispituju alate za digitalnu sigurnost pokazuju da korisnici koji aktivno koriste značajke blokiranja i prijavljivanja imaju manju izloženost ponovljenom uznemiravanju u usporedbi s onima koji ne koriste te mehanizme (Hinduja i Patchin, 2014).

Upravljanje platformama, regulacija i politike digitalne sigurnosti

Iako tehnologije kibernetičke sigurnosti pomažu u otkrivanju nasilnog ponašanja, šire strukture upravljanja digitalnim platformama također igraju ključnu ulogu u sprječavanju kibernetičkog nasilja. Upravljanje platformama odnosi se na pravila, politike i mehanizme provedbe putem kojih online platforme reguliraju ponašanje korisnika i održavaju sigurna digitalna okruženja. Ovi sustavi upravljanja uključuju smjernice zajednice, politike moderiranja, postupke prijavljivanja i disciplinske mjere poput uklanjanja sadržaja ili suspenzije računa. Znanstvenici koji proučavaju regulaciju digitalnih platformi naglašavaju da štetno online ponašanje često traje ne samo zbog tehnoloških ograničenja već i zbog nejasnih ili nedosljedno provedenih politika (Gillespie, 2018). Učinkovito upravljanje stoga zahtijeva i dobro definirana pravila i transparentne mehanizme provedbe. Standardi zajednice predstavljaju primarni regulatorni okvir putem kojeg platforme pokušavaju kontrolirati nasilnu komunikaciju. Velike tvrtke društvenih medija objavljuju detaljne smjernice u kojima se navodi koji su oblici ponašanja zabranjeni, uključujući uznemiravanje, prijetnje, lažno predstavljanje i otkrivanje privatnih informacija. Ove politike funkcioniraju kao digitalni ekvivalenti normi ponašanja u fizičkim zajednicama. Istraživanja koja ispituju provedbu smjernica zajednice pokazala su da jasno definirana pravila mogu obeshrabriti uznemiravanje kada korisnici razumiju posljedice njihovog kršenja (Matias, 2019). Međutim, provedba ostaje izazovna jer digitalne platforme svakodnevno ugošćuju milijune interakcija, što otežava praćenje svakog slučaja zlostavljanja.

Sustavi za moderiranje sadržaja stoga su središnja komponenta upravljanja platformom. Moderiranje uključuje pregled i uklanjanje štetnog sadržaja, istraživanje korisničkih prijava i primjenu sankcija kada je to potrebno. Moderiranje mogu provoditi automatizirani sustavi, profesionalni moderatorski timovi ili volonteri zajednice, ovisno o strukturi platforme. Studije praksi moderiranja pokazuju da je hibridni pristup koji kombinira automatizirano filtriranje i ljudsku prosudbu najučinkovitiji.

Automatizirani sustavi mogu brzo otkriti sumnjive obrasce sadržaja, dok ljudski moderatori mogu procijeniti kontekst i namjeru (Roberts, 2019). Na primjer, moderatori često moraju utvrditi predstavlja li poruka uznemiravanje, satiru ili legitimnu kritiku, što zahtijeva kulturno i jezično razumijevanje koje automatiziranim sustavima možda nedostaje. Posljednjih godina nacionalne vlade i međunarodne organizacije također su počele aktivnije regulirati online platforme. Regulatorni okviri pokušavaju definirati odgovornosti pružatelja digitalnih usluga u rješavanju štetnog online sadržaja. Jedan značajan primjer je Zakon o digitalnim uslugama Europske unije (DSA), koji utvrđuje obveze za velike online platforme da identificiraju rizike povezane s ilegalnim ili štetnim sadržajem i da provedu transparentne postupke moderiranja. Uredba zahtijeva od tvrtki da osiguraju pristupačne mehanizme izvještavanja, objasne odluke o moderiranju i smanje sistemske rizike poput uznemiravanja ili koordiniranih kampanja zlostavljanja (Europska komisija, 2022). Nametanjem zahtjeva za odgovornost, takve politike imaju za cilj osigurati da digitalne platforme aktivno rješavaju štetno online ponašanje, umjesto da regulaciju u potpunosti prepuste dobrovoljnim internim politikama. Još jedan važan mehanizam upravljanja uključuje mjere transparentnosti i odgovornosti. Platforme sve više objavljuju izvješća o transparentnosti koja opisuju koliko je štetnog sadržaja uklonjeno, koliko je računa suspendirano i koje su vrste kršenja bile najčešće. Ova izvješća omogućuju istraživačima i regulatorima da procijene provode li se politike moderiranja učinkovito. Studije o praksama transparentnosti platformi pokazuju da javno izvještavanje potiče platforme na poboljšanje provedbe jer tvrtke postaju odgovornije prema korisnicima, istraživačima i kreatorima politika (Suzor, 2019). Transparentnost također pomaže u prepoznavanju obrazaca zlostavljanja, omogućujući istraživačima da bolje razumiju kako se kibernetičko nasilje širi digitalnim zajednicama. Strategije upravljanja također uključuju modele moderiranja temeljene na zajednici, gdje sami korisnici sudjeluju u održavanju sigurnosti platforme. Neki online forumi omogućuju iskusnim korisnicima da prijave štetan sadržaj ili privremeno ograniče zlostavljačke račune dok moderatori ne pregledaju situaciju. Istraživanje upravljanja online zajednicom pokazuje da sudjelovanje korisnika može ojačati provedbu jer članovi zajednice često prepoznaju štetno ponašanje brže od centraliziranih timova za moderiranje (Lampe & Resnick, 2004). Međutim, moderiranje zajednice mora biti pažljivo strukturirano kako bi se izbjegla pristranost, zlouporaba ovlasti ili nepravedno ciljanje određenih korisnika.

Zaključak

Kibernetičko nasilje pojavilo se kao jedan od najznačajnijih oblika štetnog ponašanja, odražavajući složenu interakciju između tehnoloških sustava i ljudske društvene dinamike. Za razliku od tradicionalnog nasilja, kibernetičko nasilje događa se unutar online mreža gdje komunikacija može biti anonimna, brzo distribuirana i trajno snimljena. Ove karakteristike pojačavaju utjecaj uznemiravanja i otežavaju žrtvama bijeg od štetnih interakcija. Kako se platforme digitalne komunikacije nastavljaju širiti, razumijevanje mehanizama koji omogućuju kibernetičko nasilje postaje sve važnije za razvoj učinkovitih strategija prevencije. Kibernetičko nasilje se ne može shvatiti isključivo kao bihevioralni ili psihološki problem, već je ono je usko povezano sa strukturnim osobinama digitalnih platformi. Mehanizmi poput algoritamskog pojačavanja i upornog digitalnog sadržaja stvaraju uvjete koji omogućuju brzo širenje uznemiravanja i dosezanje velike publike. Psihološka istraživanja o online ponašanju pokazuju da ove tehnološke karakteristike mogu smanjiti društvenu odgovornost i povećati agresivnu komunikaciju. Kao rezultat toga, rješavanje kibernetičkog nasilja zahtijeva ispitivanje i ponašanja korisnika i tehnoloških okruženja u kojima se te interakcije događaju. Mehanizmi kibernetičke sigurnosti predstavljaju važnu komponentu ovog odgovora. Automatizirani sustavi moderiranja, alati za praćenje ponašanja, infrastruktura za izvještavanje i sustavi za upravljanje identitetom omogućuju platformama da učinkovitije otkriju i ograniče štetnu komunikaciju. Ove tehnološke mjere omogućuju platformama da ranije interveniraju u incidentima uznemiravanja i

smanje vidljivost zlostavljačkog sadržaja. Istovremeno, politike upravljanja i propisi platforme igraju ključnu ulogu u određivanju koliko se dosljedno primjenjuju ovi alati i koliko se ozbiljno rješava štetno ponašanje. Međutim, sama tehnologija ne može u potpunosti eliminirati kibernetičko nasilje. Upornost online uznemiravanja također odražava društvene norme, grupnu dinamiku i načine na koje pojedinci komuniciraju unutar digitalnih zajednica. Iz tog razloga, učinkovita rješenja moraju integrirati alate za kibernetičku sigurnost sa širim društvenim strategijama koje potiču odgovorno online ponašanje i jačaju digitalnu odgovornost. Kada tehnološke zaštitne mjere, upravljanje platformom i svijest korisnika djeluju zajedno, digitalna okruženja postaju otpornija na štetne interakcije. Kibernetičko nasilje treba shvatiti i kao tehnološki i kao društveni izazov unutar modernih digitalnih sustava. Njegovo sprječavanje zahtijeva koordinirane napore programera platformi, kreatora politika, edukatora i samih korisnika. Kombiniranjem inovacija u kibernetičkoj sigurnosti s odgovornim upravljanjem i sviješću zajednice moguće je stvoriti sigurnija digitalna okruženja koja podržavaju komunikaciju i suradnju bez uznemiravanja i zlostavljanja.

Literatura:

1. Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You can't stay here: The efficacy of Reddit's 2015 ban examined through hate speech communities. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1–22.
2. Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 512–515.
3. European Commission. (2022). The Digital Services Act: Ensuring a safe and accountable online environment. Publications Office of the European Union, Brussels.
4. Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media (pp. 68–97). Yale University Press, New Haven.
5. Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media (pp. 107–114). Yale University Press, New Haven.
6. Hinduja, S., & Patchin, J. W. (2014). Bullying beyond the schoolyard: Preventing and responding to cyberbullying (2nd ed., pp. 118–121). Sage Publications, Thousand Oaks, CA.
7. Jhaver, S., Ghoshal, S., Bruckman, A., & Gilbert, E. (2019). Online harassment and content moderation: The case of blocking on social media. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–27.
8. Kowalski, R. M., Giumetti, G. W., Schroeder, A., & Lattanner, M. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research. Psychological Bulletin, 140(4), 1073–1137.
9. Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). Cyberbullying: Bullying in the digital age (2nd ed., pp. 44–47). Wiley-Blackwell, Malden, MA.
10. Lampe, C., & Resnick, P. (2004). Slash(dot) and burn: Distributed moderation in a large online conversation space. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 543–550.
11. Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in online communities. Proceedings of the National Academy of Sciences, 116(20), 9785–9789.
12. Patchin, J. W., & Hinduja, S. (2015). Bullying beyond the schoolyard: Preventing and responding to cyberbullying (2nd ed., pp. 23–28). Sage Publications, Thousand Oaks, CA.

13. Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media (pp. 58–63). Yale University Press, New Haven.
14. Slonje, R., Smith, P. K., & Frisén, A. (2013). The nature of cyberbullying and strategies for prevention. Computers in Human Behavior, 29(1), 26–32.
15. Suler, J. (2004). The online disinhibition effect. CyberPsychology & Behavior, 7(3), 321–326.
16. Suzor, N. (2019). Lawless: The secret rules that govern our digital lives (pp. 120–125; 175–179). Cambridge University Press, Cambridge.
17. Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other (pp. 180–186). Basic Books, New York.
18. Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), The SAGE handbook of interpersonal communication (4th ed., pp. 443–479). Sage Publications, Thousand Oaks, CA.

SECURITY AND CYBERBULLYING

Dr. Mirko Goreta, Senior Lecturer

Marko Marulić University of Applied Sciences in Knin

Abstract: *Cyberbullying has become a widespread problem in modern digital communication environments. Social media platforms, messaging applications and online communities enable individuals to interact via large networks, but these same technologies can also enable harassment, intimidation and persistent online bullying. Unlike traditional bullying, cyberbullying is often amplified by the characteristics of digital platforms, including anonymity, rapid information sharing and permanent visibility of content. This paper examines cyberbullying from the perspective of digital security and online platform design. It analyses the mechanisms through which cyberbullying develops in digital environments, the psychological and sociological factors that influence aggressive online behaviour and the technological systems used to detect harmful interactions. Special attention is paid to cybersecurity mechanisms such as automated moderation systems, behaviour monitoring tools and platform governance structures that aim to reduce harassment in online spaces. The analysis shows that cyberbullying arises from the interaction between user behavior and the technological architecture of digital platforms. Addressing this problem therefore requires an integrated approach that combines security technologies, responsible platform governance, and greater awareness of digital communication practices. Strengthening these elements can contribute to safer online environments where harmful behavior is more effectively detected, limited, and prevented.*

Keywords: *cyberbullying, cybersecurity, digital platforms, content moderation, platform governance.*