

Pregledni znanstveni članek/Article (1.02)

Bogoslovni vestnik/Theological Quarterly 85 (2025) 4, 845—860

Besedilo prejeto/Received:08/2025; sprejeto/Accepted:09/2025

UDK/UDC: 179.8:004.8:316.614

DOI: 10.34291/BV2025/04/Centa

© 2025 Centa Strahovnik et al., CC BY 4.0

Mateja Centa Strahovnik in Vojko Strahovnik
**Sovražnost in nasilje v digitalni kulturi
 ter dobi umetne inteligence**
*Hate and Violence in the Digital Culture
 and the Era of AI*

Povzetek: Prispevek¹ obravnava oblike, značilnosti in posebnosti spletnega sovraštva, sovražnosti in nasilja v digitalni kulturi. Hkrati so izpostavljene nekatere oblike, ki so posebej značilne za umetno inteligenco oz. ki jih uporaba teh tehnologij bistveno predrugači. Prispevek se ukvarja tudi s temeljnimi značilnostmi takšnega sovraštva in nasilja ter še posebej izpostavlja tiste, ki zadevajo čustveno oz. afektivno domeno. Kot ena izmed podmen za preseganje spletnega sovraštva in nasilja je obravnavano sočutje. Zaključek je posvečen vprašanju ustreznega oz. celostnega razumevanja sočutja in njegove vloge v digitalni kulturi, vključno z vidiki, ki jih izpostavlja teologija. Sočutje ni le moralno pomembno in zasebno čustvo, temveč družbeno in politično pomembna naravnost, ki jo je treba gojiti, da bi v digitalni kulturi krepili razvoj pravične družbe, celostno spoštovanje človekovega dostojanstva in ustrezno delovanje demokratičnih institucij.

Ključne besede: spletno sovraštvo, sovražnost, nasilje, digitalna kultura, umetna inteligenca

Abstract: The article discusses the forms, characteristics and specificities of online hate, hostility and violence in the digital culture. At the same time, it highlights some of the forms that are specific to artificial intelligence, or that are significantly transformed by the use of AI technologies. The paper also deals with the fundamental characteristics of such hatred and violence, highlighting in particular those that concern the emotional or affective domain. The article discusses compassion as one of the subdomains for overcoming online hate and violence. In conclusion, the article addresses the question of an adequate or holistic understanding of compassion and its role in digital culture, including the aspects highlighted by theology. Compassion is not only a morally central and private emotion,

¹ Prispevek je nastal v okviru zagonskega raziskovalnega programa „Presečišče kreposti, izkustva in digitalne kulture: Etični in teološki uvidi“, ki ga financira Univerza v Ljubljani, raziskovalnega programa „Religija, etika, edukacija in izzivi sodobne družbe (P6-0269)“ in temeljnega raziskovalnega projekta „Teologija, digitalna kultura in izzivi na človeka osredičene umetne inteligence (J6-4626)“, ki ju sofinancira Javna agencija za raziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS). Za podporo tej raziskavi niso bili ustvarjeni nobeni novi podatki.

but a socially and politically important attitude that needs to be cultivated in the digital culture in order to support the development of a just society, the full respect for human dignity and the proper functioning of democratic institutions.

Key Words: online hate, hostility, violence, digital culture, artificial intelligence

1. Uvod

V digitalni dobi, ki jo vse bolj zaznamujeta globalna povezanost in tehnično-znanstveni napredek na področju digitalnih tehnologij, se za oblikovanje in širjenje sovraštva in sovražnosti ter nasilja vzpostavlja nov okvir oz. kontekst. Vanj spada npr. spletni sovražni govor, digitalno nasilje in algoritmično pogojene oblike diskriminacije, ki presegaajo tradicionalne pojme sovražnosti in so v pomembni meri pogojene z razvojem umetne inteligence (UI) ter drugih digitalnih tehnologij. Namen prispevka je razčleniti pojavne oblike sovraštva, sovražnosti in nasilja, kot se manifestirajo v digitalnih okoljih, ter raziskati možnosti njihovega preseganja. Posebna pozornost bo posvečena čustvenim oz. afektivnim razsežnostim – tako pri sovražnostih samih kot pri njihovem preseganju.

V prispevku po uvodnem razdelku sledi razdelek, ki obravnava oblike, značilnosti in posebnosti spletnega sovraštva, sovražnosti in nasilja. Hkrati so izpostavljene nekatere oblike, ki so značilne zlasti za UI oz. ki jih uporaba tehnologij UI bistveno predrugaja. Posebej so izpostavljene oblike sovražnosti in nasilja do religije. Ob koncu tega razdelka pa se prispevek ukvarja še s temeljnimi značilnostmi takšnega sovraštva in nasilja in še posebej izpostavlja tiste, ki zadevajo čustveno oz. afektivno domeno. V tretjem razdelku je obravnavano sočutje kot ena izmed podmen za preseganje spletnega sovraštva in nasilja. V zaključku pa je obravnavano vprašanje ustreznega oz. celostnega razumevanja sočutja in njegove vloge v digitalni kulturi – vključno z vidiki, ki jih izpostavlja teologija. Sočutje ni le moralno pomembno in zasebno čustvo, temveč družbeno in politično pomembna naravnost, ki jo je treba gojiti, da bi v digitalni kulturi krepili razvoj pravične družbe, polno spoštovanje človekovega dostojanstva in ustrezno delovanje demokratičnih institucij.

2. Oblike, značilnosti in posebnosti spletnega sovraštva, sovražnosti in nasilja z vidika digitalne kulture ter umetne inteligence

Digitalno kulturo lahko opredelimo kot skupek individualnih in kolektivnih prepričanj, vrednot in praks, ki so nastale zaradi vsesplošne uporabe in integracije digitalnih tehnologij na večini področij našega življenja. V ožjem smislu zajema način, kako ljudje sodelujejo, komunicirajo, ustvarjajo in uporabljajo informacije v digitalnem svetu, v širšem smislu pa zadeva tudi druge pomembne vidike človeškega življenja, kot so odnosi, izkušnje, učenje, ideali, kultura, duhovnost, ustvarjalnost. Digitalno kulturo oblikujejo predvsem svetovni splet, družbeni mediji, digitalne naprave in spletne skupnosti (Centa Strahovnik 2025, 133).

V tej novi kulturi pa sta prisotna tudi sovraštvo in nasilje, ki sta zaznamovana z določenimi značilnostmi in posebnostmi tega okolja. O spletnem nasilju in sovraštvu najpogosteje govorimo, ko posamezniki ali skupine na spletu oz. v digitalnem okolju izrabljajo svojo (pre)moč nad drugimi posamezniki ali skupinami in s svojimi dejanji občutno posegajo v njihove pravice oz. dobrobit. Gre za to, da se storilec (posameznik ali skupina) specifično usmeri na drugo osebo ali skupino, da bi ji povzročil škodo – vključno s čustveno škodo, prisilnim nadzorom ali grožnjami s fizičnim nasiljem ali z drugimi vrstami nasilja. Lahko pa gre že za takšno nasilje samo (Keats Citron 2015).

V prispevku obravnavamo tako oblike spletnega sovraštva in sovražnosti kot tudi oblike spletnega nasilja, kajti oboje je pogosto tesno prepleteno oz. se lahko do določene mere prekriva. Obstaja pa seveda nekaj osnovnih razlik. Spletna sovražnost oz. izražanje sovraštva na spletu je primarno verbalno oz. izrazno, medtem ko je nasilje povezano z dejanji (pogosto tudi kaznivimi). Namen sovraštva je najpogosteje marginalizirati, poniževati ali izzivati, medtem kot gre pri nasilju za neposredno povzročanje škode oz. strahu. Spletno sovraštvo ni nujno nekaj, kar bi bilo glede na zakonodajo kaznivo, medtem ko je spletno nasilje praviloma prepovedano in kaznivo. Je pa spletno sovraštvo nekaj, kar do nasilja lahko privede ali pa to nasilje normalizira. Še posebej lahko to pride do izraza v tistih primerih, ko so pravila in omejitve, npr. glede izražanja na družbenih omrežjih, opredeljeni zelo široko in nedoločno (Hietanen idr. 2022, 445).

2.1 Oblike oz. vrste spletnega sovraštva

Oblike spletnega nasilja in sovraštva lahko razvrstimo v posamezne vrste oz. kategorije glede na več razsežnosti oz. vidikov. Z vidika namere oz. namernosti lahko razlikujemo med namernimi in nenamernimi dejanji. Z vidika identitete storilcev lahko ločujemo med posameznimi storilci, skupinami, institucijami (v najširšem smislu) ter samimi sistemi UI oz. algoritmi. Z vidika tarč tega nasilja oz. sovražnosti lahko podobno ločujemo med posamezniki, skupinami, identitetami (ali prepletom identitet) oz. strukturami. Z vidika vidnosti tega nasilja lahko razlikujemo med oblikami, ki so vidne (običajno namenoma), in oblikami, ki so prikrite. Tem osiščem razlikovanja je blizu tudi tisto glede razširjenosti ali dostopnosti spletnih prostorov, kjer se takšno sovraštvo pojavlja – npr. javno dostopne in razširjene platforme, kot so priljubljena družbena omrežja, ali pa platforme, ki so obrobne ali celo namenoma prikrite in zasebne. Pomembno je lahko tudi razlikovanje med enkratnimi dejanji ter ponavljajočim se oz. vztrajnim delovanjem (Thomas idr. 2021).

V nadaljevanju podajamo kategorizacijo spletnega nasilja in sovražnosti, ki se v prvi vrsti opira na razlike med vrstami dejanj oz. podobnosti med dejanji v posamezni kategoriji.²

² V tej taksonomiji oblik nasilja in sovražnosti ter v nadaljnji razpravi se omejujemo na nasilje in sovražnosti, ki jih izvajajo posamezniki ali skupine posameznikov, deloma pa tudi korporacije, ki razvijajo sisteme UI, jih ponujajo v uporabo ali so lastniki družbenih omrežij. Ne obravnavamo pa oblik, katerih nosilec je država (ali nje podobne tvorbe, denimo mednarodne organizacije ipd.), kot so npr. informacijsko bojevanje in socialni inženiring (angl. *brigading; cognitive/information warfare, social engineering*) ali pa (nezakonit) množični nadzor in zbiranje podatkov (angl. *unlawful mass surveillance; dragnet*).

Spletno nadlegovanje (angl. *online harassment, cyberbullying*) in *razkrivanje informacij*. Pri prvem gre za namerno povzročanje težke čustvene stiske z vztrajnim sovražnim in negativnim spletnim govorom, ki prihaja z več strani in se izraža na več načinov (spletno ustrahovanje, žaljenje, sramotenje, namerno povzročanje občutka izključenosti, ignoriranje in izključevanje, širjenje neresnic, trolanje oz. provociranje, spodbujanje k samomoru ali samopoškodovalnemu vedenju). Takšno nadlegovanje se lahko pretvori tudi v delovanje v dejanskem svetu, npr. osebno-podatkovno razkrivanje (angl. *doxing*) kot javno razkritje osebnih podatkov s pozivom k nadlegovanju ali *swating* (angl. *swatting*) kot lažna prijava posameznika ali skupine policiji ali drugim podobnim službam, ki na podlagi lažnih prijav grobo posežejo v zasebnost nedolžne žrtve (Ybarra idr. 2006; Slonje idr. 2008).

Spletno zalezovanje (angl. *online stalking*) ima običajno ožji pomen kot nadlegovanje. Gre za spletno ravnanje, ki povzroči, da se žrtev neposredno boji za svojo (fizično) varnost. To ravnanje zajema grožnje (z nasiljem), npr. preko e-sporočil – tako žrtvam kot tudi drugim z njo povezanim osebam; laži o žrtvi in lažne oglase (npr. za spolne usluge, kjer so navedeni osebni podatki žrtve in prirejene fotografije), ki škodijo ugledu žrtve; pozive neznanecem, naj žrtev fizično poškodujejo; zlorabo zasebnosti: vdori v računalnik, telefon; fotografiranje in snemanje s spletno kamero ipd.

Spletni sovražni govor (angl. *hate speech*) zajema žalitve in druge diskriminatorne interakcije, ki merijo na določila ali lastnosti osebe (zaničevanje in poniževanje osebe zaradi osebnih značilnosti, kot so rasa, vera, spol, spolna usmerjenost, starost, zdravstveno stanje, invalidnost, izobrazba ...).

Spletno zavajanje je strategija, ki jo posamezniki ali skupine uporabljajo za manipulacijo in zavajanje ljudi, da ti razkrijejo občutljive informacije ali ravnajo na način, s katerim ogrožajo svojo varnost. Med drugim obsega pretvarjanje in izsiljevalske programe (angl. *phishing*) ter širjenje vsebin za radikalizacijo, novačenje ali spodbujanje nasilja iz ideoloških razlogov (politika, religije, aktivisti za okolje, cepiva ipd.). V ta okvir lahko sodijo dezinformacijske kampanje in zarotniške mreže, ki navadno delujejo na sistematičen in usklajen način. Najpogosteje gre za sistematično širjenje lažnih informacij, za manipuliranje s prepričanji, podžiganje sovraštva ali destabiliziranje zaupanja v institucije (Mahl idr. 2025). Kot posebno obliko nasilja lahko v ta okvir umestimo tudi različne finančne prevare, ki so usmerjene predvsem v najbolj ranljive skupine posameznikov, ter krajo identitete.

Spletna cenzura (angl. *online censorship*). Spletna cenzura, ki jo izvajajo posamezniki ali skupine, se nanaša na neformalne mehanizme zatiranja ali utišanja vsebin ali vidikov v digitalnih prostorih. V nasprotju z državno cenzuro nedržavna spletna cenzura pomeni usklajena dejanja posameznih akterjev, ki si prizadevajo urejati diskurz v skladu z določenimi normami, vrednotami ali cilji. Kot takšna lahko vključuje usklajeno prijavljanje objav ali uporabniških računov/profilov na določeni platformi, kampanje nadlegovanja in utišanja, pozivanje k preklicu ali prepovedi vsebin ali govora posameznikov, ustvarjanje poplave vsebin, ki določeno domeno ali polje razprave preplavijo z nepomembnim ali motečim gradivom. Te prakse so pogosto neformalne in brez centraliziranega nadzora, zato so nepregledne in jih je težko regulirati (Aceto 2015).

Ekstremistične vsebine in pozivi k neposrednemu nasilju. Gre za objavlanje in razširjanje ekstremističnih vsebin – npr. teroristične propagande –, ki lahko vodijo do terorističnih dejanj, ali ekstremističnih ideologij, ki spodbujajo skrajne politične, verske ali družbene ideologije s ciljem radikalizacije posameznikov. Pogosto so prisotni tudi konkretizirani pozivi k nasilju (Aly idr. 2016; Thompson 2011).

Posebno pozornost tovrstno sovraštvo in nasilje terja, ker je pogosto usmerjeno v ranljive skupine. Podatki za Slovenijo za leto 2018 tako med drugim kažejo, da je v tem letu kar 56 % učenk in 50 % učencev od 7. do 9. razreda doživelo vsaj eno izmed oblik spletnega nadlegovanja – nadlegovanje po mobilnih telefonih (sporočila, klici, videoposnetki, fotografije), na spletu (elektronska pošta, spletna sporočila, klepetalnice, spletne strani) ali z drugimi tehnologijami. Gre za nadlegovanje, kjer kdo drugega žali, nadleguje, zasmehuje, ogroža in ustrahuje. Med srednješolci je tako izkušnjo doživelo kar 65 % dijakinj in 55 % dijakov (Završnik in Hooper 2025, 5).

2.2 Oblike spletnega sovraštva, sovražnosti in nasilja, ki jih omogoča oz. krepi raba UI

V tem razdelku se bomo podrobneje posvetili oblikam, ki so tesno povezane s tehnologijo oz. sistemi UI, tj. oblikam, ki jih ti sistemi omogočajo, predružačijo oz. celo sami tvorijo. Z razvojem sistemov UI, zlasti velikih jezikovnih modelov, ki v določenih ozirih postajajo vedno bolj podobni človeku oz. vidikom njegovega vedenja, so se nekatere oblike sovraštva in nasilja predružačile, okrepile oz. so nastale nove. Gre za sisteme, ki uporabljajo ogromne količine podatkov, da se učijo in samostojno tvorijo vsebine, ki jih lahko potem tudi samodejno širijo ter profilirajo in izbirajo naslovnike teh vsebin (Strahovnik 2023). Eden izmed neposredno pomembnih vidikov sistemov UI je ta, da lahko kot sredstvo krepijo namene oz. izvajajo naloge, ki jim jih dodeljuje človek. Pri tem so zelo učinkoviti – npr. pridobivanje obsežnega nabora podatkov, hitrejša analiza podatkov, manipulacija vsebin (globoki ponaredki), množično širjenje vsebin (Steffensen 2020). Po drugi strani pa lahko privedejo do svojstvenih oblik nasilja in sovražnosti, ki z ukazi ali nameni ljudi niso nujno neposredno povezane – npr. algoritemsko nasilje, žalitve, ki jih tvori UI ipd. (Strahovnik 2024). V nadaljevanju navajamo nekatere temeljne vrste oz. oblike nasilja in sovražnosti, ki so tesno povezane s sistemi UI.

Globoki ponaredek (angl. *deep fake*) je spletna vsebina, ki je s pomočjo posebnih algoritmov UI spremenjena ali ustvarjena tako prepričljivo, da jo težko prepoznamo kot neresnično. Takšna oblika manipulacije vsebine je sicer bila mogoča že prej, a široko dostopni sistemi UI sedaj omogočajo množično ustvarjanje globokih ponaredkov, saj je tehnologija dostopna vsakemu, ki ima dostop do spleta. Globoki ponaredki sami po sebi še niso oblika sovražnosti ali nasilja, lahko pa to postanejo, če gre za vsebine, ki so žaljive ali manipulativne – in se uporabljajo z namenom npr. izsiljevanja, groženj, nadlegovanja, razširjanja propagande (Maras idr. 2018; Westerlund 2019).

Tovarne memov (angl. *meme factories*). Mem je sporočilo oz. jedrna ideja, zlasti v obliki slike z napisom ali kratkega posnetka, ki se tako prenaša v elektronski komunikaciji in služi kot referenca, ki je navzven ne prepoznamo nujno kot takšne

(npr. žabec Pepe kot simbol ideologije nadvlade in večvrednosti bele rase). Sistemi UI omogočajo samodejno tvorjenje in razširjanje takšnih vsebin, ki predstavljajo nasilje ali sovražnost, kadar gre za vsebine, ki so nestrpne, izključujoče, ponižujoče ipd., ali pa vsebine, ki pozivajo k nasilju (Abidin 2020).

UI pogovorni roboti (angl. *chatbots*). Gre za pogovorne sisteme (npr. Nomi, Grok), ki temeljijo na UI – natančneje, na velikih jezikovnih modelih. Kadar pri njihovi implementaciji upravljalci sledijo izhodiščni filozofiji svobode govora (vse je dovoljeno – brez omejitev), lahko samodejno ustvarjajo različne žaljivke in spodbujajo sovraštvo. Lahko pa delujejo v vlogi t. i. *trolbotov*, katerih namen je s pomočjo žaljivega in sovražnega ter razdiralnega govora vzbujaati čustvene odzive kar najširšega kroga ljudi ali pa posameznike spodbujati k nasilju do drugih ali samih sebe (samomori, samopoškodovanje) (Adamopoulou idr. 2020; Patel idr. 2024; Fire idr. 2025).

Samodejno generiran sovražni govor oz. žaljivke. Generativni sistemi UI vzpostavljajo nove razsežnosti problematike žaljivk in sovražnega govora v digitalnih okoljih, saj lahko takšne vsebine ustvarjajo samostojno, kar sproža nove pomisleke glede odgovornosti, pristranskosti in krepitev škodljivih ideologij. Te žaljivke (angl. *slurs*) so usmerjene proti posameznikom ali skupinam. Tako tvorjen sovražni govor zajema tudi izraze, ki spodbujajo sovraštvo, diskriminacijo, sovražnost ali nasilje do določenih skupin. To lahko vključuje tako odkrito agresivne vsebine kot tudi bolj prikrite oblike žaljenja, zaničevanja, poniževanja ali stereotipizacije (Sarkar 2025).

Profiliranje s pomočjo UI. Profiliranje z UI pomeni uporabo algoritemskih sistemov za sklepanje, razvrščanje ali napovedovanje posameznikovih ali skupinskih lastnosti, vedenja ali izbire na podlagi vhodnih podatkov (denimo pri zagotavljanju varnosti, priseljevalni politiki, zaposlovanju, trženju, izobraževanju ipd.). Čeprav se profiliranje z UI pogosto prikazuje kot koristno – npr. v smislu učinkovitosti in personalizacije – in torej kot nekaj pozitivnega, je lahko tudi sovražno, škodljivo ali celo nasilno (izključevanje, prepoved oz. ukinitve svobode govora in izražanja, zaničevanje določenih perspektiv), zlasti kadar se opira na obstoječe predsodke in neenakosti.

Algoritična diskriminacija (angl. *algorithmic bias*): sistemi za nadzor, kot so denimo programi za prepoznavanje obrazov in drugi sistemi, so lahko sistematično pristranski (rasna, spolna in starostna pristranskost). Do pristranskosti pride, ko pristranski učni podatki ali sistematične napake v algoritmih strojnega učenja povzročijo nepoštene, diskriminatorne rezultate, kar pogosto odraža ali krepi obstoječe socialno-ekonomske, rasne in spolne pristranskosti.

2.3 Sovražnost do religij(e)

V tem razdelku izpostavljamo posebno obliko sovražnosti in nasilja z vidika vsebine oz. religije kot tarče tega nasilja. Religije, religijske skupnosti in posamezniki, ki jim pripadajo, so lahko žrtev katere koli zgoraj omenjene oblike – vključno z oblikami spletnega sovraštva, sovražnosti, nasilja, ki jih omogoča UI (npr. algoritemska pristranskost zoper pripadnike določene veroizpovedi ali spletne cenzure določenih verskih vsebin). Sovražnost in nasilje sta v tem smislu usmerjena proti določenim verskim skupnostim – vključujeta kritiko verskih prepričanj in napade na same vernike (Bermudez-Villalva

idr. 2025). Najpogosteje so te oblike usmerjene na tri glavne monoteistične religije. Taksonomije spletnega sovraštva v zvezi z religijami razlikujejo med štirimi osnovnimi kategorijami: sovražnost (kot močno izraženo sovraštvo in zavračanje), žalitve in razčlovečenje (neposredni napadi, žaljivke in žaljivi izrazi, ki se lahko nanašajo tudi na skupino), grozeč govor (z namero, željo povzročiti posamezniku ali skupini škodo, kar vključuje upravičevanje povzročene škode oz. podporo takemu ravnanju) ter žaljiv in zlorabljač humor (implicitna sovražnost, ki se zakriva s šalo, oz. šale, ki vzdržujejo in razširjajo negativne stereotipe) (Ramponi et al. 2022, 6–7). Večina teh oblik uporablja retorične prijeme in druga izrazna sredstva, ki jih samodejni sistemi za prepoznavo spletnega sovražnega govora ne morejo zlahka zaznati.

Kaj so osrednje posebnosti spletnega sovraštva, povezanega z religijo? Najprej to, da tarče takšnega sovraštva niso zgolj posamezniki ali skupnosti, ampak tudi pričanja, doktrine, sveti simboli in ikone, obredne prakse, verske stavbe – vključno z njihovo skrunitvijo – ipd., kar je pogosto težko preprečevati ali preganjati. Posebno pogosta oblika je, da dejanske skrunitve (npr. zažiganje svetih spisov ali rušenje stavb) s pomočjo digitalnih medijev presežejo svoje lokalne okvire in kontekst ter sprožajo valove odzivov, tudi nasilnih. Verska sovražnost pogosto vključuje zanikanje svetega (bogokletje) ali nadzorovanje mejá (oznake krivoverca/odpadnika); pri tem se digitalna oz. spletna mobilizacija lahko hitro stopnjuje. Prisotna je izrazita občutljivost na dogodke in koledar, saj se te oblike sovražnosti in nasilja okrepijo ob sprožilnih dogodkih (teroristični napadi, geopolitični pretresi), v času verskih praznikov in spominskih dni. Ker je polje religije in religijskih prepričanj pogosto na meji pravno zaščitene govora in zaščitene skupin, prihaja pri preganjanju takšnega sovraštva do neskladij in odprtih vprašanj, ki jih različne zakonodaje in oblasti tudi obravnavajo različno. In nenazadnje – napake in zmote algoritmov UI so, ko gre za religijo, toliko bolj verjetne. Klasifikatorji, ki so usposobljeni za prepoznavo splošnih oz. običajnih nevarnosti ali toksičnosti, kodirane verske žalitve običajno težje zaznajo, saj je fenomena religije pogosto ni mogoče ujeti v enostavne pojmovne sheme in preračunske postopke (Ramponi et al. 2022; Mostafazadeh Davani idr. 2023).

Zanimivo pri tej vrsti sovražnosti je, da je slabo raziskana – čeprav gre za pomemben družbeni problem oz. vprašanje. Ramponi in soavtorji (2022) izpostavljajo, da raziskave, ki se ukvarjajo s sovražnim govorom na spletu, v nasprotju z raziskavami drugih vrst sovražnega govora, kot je npr. zaznavanje rasizma, sovraštva na podlagi spola idr. (kar je raziskano kar podrobno), verski sovražni govor na spletu bolj ali manj izpuščajo. Posebej povečana nevarnost za sovražni govor je v slabo moderiranih in nadzorovanih digitalnih prostorih, ki prispevajo k širjenju sovražnega govora. Na podlagi podatkov, pridobljenih s spletne platforme 4chan, ki je znana kot nedomoderirano spletno okolje – kjer se v praksi izvaja oz. je dovoljena neomejena svoboda govora –, dobimo vpogled v dinamiko sovražnega govora. Najbolj razširjen je rasizem, ki predstavlja 35,92 % sovražnih objav oz. vsebin. Na drugem mestu je religijsko pogojeni sovražni govor (23,36 %), sledi sovražni govor na podlagi spolne usmerjenosti (16,49 %), nato pa seksizem (12,01 %), sovražni govor do oseb z ovirami (1,73 %) in drugo (10,49 %) (Bermudez-Villalva et al. 2025). Iz raziskave lahko vidimo, da gre pri verskem sovražnem govoru za zelo resen problem, saj je po pogostosti na

drugem mestu. Hkrati porast takšnega sovražnega govora običajno vodi tudi do porasta sprožilnih dogodkov, ki vodijo v nasilje.

Verski sovražni govor po nekaterih študijah lahko vodi tudi do radikalizacije posameznikov, če ti niso izpostavljeni informacijam, ki bi te ideje in prepričanja izpodbijale (Ramponi idr. 2022; Thompson 2011). Zato je pomembno, da bolje razumemo, kako se verski sovražni govor na spletu izraža, katere oblike žaljivega jezika se pri tej vrsti ciljnih skupin običajno uporabljajo – in ali med različnimi religijami obstajajo razlike. To bi pripomoglo ne le k boljšemu razumevanju pojava, ampak bi omogočilo tudi razvoj in vzpostavitev učinkovitih modelov in sistemov za zaznavanje in zoperstavljanje sovražnemu govoru (Bermudez-Villalva et al. 2025; Ramponi idr. 2022) in širše reševanje sovražnosti do religij(e).

2.4 Značilnosti spletnega sovraštva, sovražnosti in nasilja

Glede digitalnega sveta oz. dobe digitalne kulture lahko izpostavimo nekaj ključnih lastnosti. Prva je razširjena možnost za skoraj popolno anonimnost, kar uporabnikom omogoča po eni strani varno in zasebno komunikacijo, po drugi pa številne zlorabe. Drugič, spletno okolje omogoča delovanje na daljavo v obsegu, ki doslej še ni bil poznan. Tretja značilnost je viralnost oz. hitro stopnjevanje in širjenje informacij, kar omogoča globalen doseg v skoraj neznatnem času. Četrto, digitalni zapis oz. odtis je praktično nemogoče izbrisati, kar pomeni, da so ti odtisi trajni. Petič, algoritmi oz. sistemi UI lahko te procese še dodatno okrepijo, saj širjenje vsebin optimizirajo in pospešujejo ter zagotavljajo, da dosežejo najširše možno občinstvo ali pa najbolj relevantne in ranljive populacije. Šestič, pred tovrstnim nasiljem se je pogosto težko neposredno braniti (Aiken 2016; Keats Citron 2015).

Hkrati lahko vse (ali večino) teh značilnosti prenesemo tudi na področje nasilja in sovražnosti. Te izpostavljene značilnosti digitalnega okolja – anonimnost, delovanje na daljavo, viralnost, trajnost digitalnih sledi in algoritemski okrepitevi – skupaj tvorijo nekakšen družbeno-tehnični ekosistem, ki oblike digitalnega sovraštva, sovražnosti in nasilja omogoča in krepi. Ko se te značilnosti prenesejo na spletno nasilje, ne razkrivajo le njegovih strukturnih dejavnikov, temveč pojasnjujejo tudi, zakaj se digitalne oblike od podobnih oblik nasilja, ki se pojavljajo tudi nasploh, kvalitativno in kvantitativno razlikujejo. Tako lahko možnost skoraj popolne anonimnosti ranljive posameznike po eni strani ščiti, vendar po drugi strani omogoča, da naročniki ali izvajalci sovražnega govora, spletnega ustrahovanja in usklajenih kampanj nadlegovanja delujejo brez strahu pred identifikacijo ali pravnimi posledicami. Ta anonimnost opogumlja sovražnost oz. njeno izražanje, spodbuja nekaznovanje in zamegljuje odgovornost. Delovanje na daljavo spletnemu nasilju omogoča, da presega geografske in časovne meje – žrtve in storilci se morda nikoli ne srečajo fizično, vendar je lahko učinek nasilja in izražanja sovraštva globoko oseben in takojšen, hkrati pa se lahko nasilje izvaja v velikem obsegu. Poleg tega so lahko kampanje, ki temeljijo na sovraštvu, asinhrono in se nadaljujejo še dolgo po tem, ko se njihovi pobudniki vanje prenehajo vključevati, kar lahko ustvarja valove zlorab. Viralnost omogoča, da se lahko škodljive vsebine lahko razširijo in v nekaj sekundah dosežejo množično občinstvo. Sovražni govor, žaljive vsebine ali prirejene slike lahko postanejo viralne in povečajo

svoj doseg – in to ne nujno zaradi svoje navidezne resničnosti ali pomembnosti, temveč zaradi čustvenih provokacij, radovednosti ali senzacionalnosti. Hkrati viralnost pomeni tudi določeno slavo, kar spodbuja posnemanje oz. mimikrijo, ki potem krepí sovražno digitalno kulturo. Trajnost in neizbrisljivost digitalnega odtisa omogoča, da se denimo žaljive obtožbe ali rasistične žaljivke na platformah in domenah ohranjajo brez časovnih mej, kar pri žrtvah lahko povzroči retravmatizacijo in občutek brezizhodnosti. In nenazadnje, algoritmi UI lahko na podlagi analize podatkov in nastopanja v vlogi priporočilnih sistemov dajejo prednost polarizirajočim, čustveno nabitim ali ekstremnim vsebinam, s čimer nehote krepíjo sovražstvo in sovražnost – po drugi strani pa t. i. mikrousmerjanje omogoča, da škodljive vsebine dosežejo čustveno ali ideološko najbolj dovzetne posameznike (Aiken 2016).

2.5 Posebnosti digitalne kulture in spletnega sovražstva, sovražnosti, nasilja

V tem zadnjem podrazdelku najprej obravnavamo čustveno zaznamovanost digitalne kulture; nato izpostavljamo, kakšne čustvene razsežnosti in čustvene posledice so povezane z spletnim nasiljem, sovražstvom in sovražnostmi – še posebej do religij.

Ena izmed posebnosti digitalne kulture je, da je – kljub temu, da gre v prvi vrsti za vidik tehnologij in njihove uporabe – izrazito čustvena (Marino idr. 2024). To velja predvsem za spletna okolja, kot npr. družbena omrežja, pogovorne robote, aplikacije za sporočila, kjer komunikacija poteka pretežno na čustveni ravni, medtem ko so druge vsebine potisnjene v ozadje. Družbena okolja se uporabljajo za samoizražanje (izražanje identitete), izražanje pripadnosti določeni skupini (trendi), pri čemer so pogosti čustveni simboli (digitalni jezik, ki vključuje emodžije, všečke, ekspresivno rabo ločil (npr. !!!, !?!, <3, asdfghjkl!!!!), gife, meme in krajše videoposnetke, ki praviloma niso dokumentarni, ampak čustveni in ekspresivni). Skupnosti se oblikujejo tudi okoli teh skupnih oz. skupinskih čustev, kot so npr. ponos, jeza in ogorčenje – zlasti pri kulturi izključevanja, črtanja oz. utišanja (angl. *cancel culture*) –, sočutje in solidarnost (npr. spletne akcije zbiranja sredstev), strah in tesnoba (skupnosti teoretikov zarot) ipd. Ta čustvena zaznamovanost zajema široko polje – od zelo površinskih pojavov, kot so všečkanje, deljenje, sledenje, strinjanje, do bolj globinskih, manj transparentnih, ko npr. UI samodejno generira lažne profile, ustvarja navidezne všečke in sledilce ali pa meme, glede na katere se posameznik identificira.

Posebnost in problem je, da algoritmi spletnih okolij, kot so Instagram, TikTok ali Twitter/X, dajejo prednost vsebini, ki vzbuja močne čustvene odzive, kot je jeza (ogorčenje), veselje (norčevanje, zasmehovanje) in strah (groza), saj to omogoča več klikov, delitev in časa preživetega na spletu, kar so bistveni vidiki za oglaševanje. Čustveno nabite objave se širijo hitreje kot nevtralne. Spletno okolje tako na neki način nagraduje čustveno izražanje, ne pa navajanje dejstev in argumentov. Algoritmi krepíjo čustveno nabito vsebino, saj posledično to prinaša večji dohodek in moč (npr. v obliki razširjenosti in prepoznavnosti platforme).

Nadalje, ta okolja lahko (s pomočjo algoritmov) spodbujajo napačno dojetje čustev, s čimer med posamezniki in skupinami širijo nestrpnost in nasilje. Ker po-

samezniki (vključno s političnimi voditelji) po spletnih družbenih omrežjih komunicirajo vse pogosteje, je to dinamiko zaznavanja čustev na spletu pomembno razumeti. Uporabniki tovrstnih omrežij mdr. pretirano zaznavajo raven moralnega ogorčenja, ki ga čutijo posamezniki in skupine, kar še dodatno neupravičeno povečuje prepričanja o medskupinski sovražnosti (Brady idr. 2023).

Posebnost tega okolja, ki hkrati predstavlja tudi problem, so še družabništvu (kakor tudi terapiji, duhovnemu in verskemu vodenju ipd.) namenjeni pogovorni roboti, saj ti vsebujejo algoritme, ki vključujejo čustvene odzive. Ti boti se prilagajajo uporabniku glede na podatke in informacije, ki jih ta vnaša in do katerih lahko algoritmi nato dostopajo tudi sami. Pogovorni robot uporabnika analizira, profilira, mu nudi personaliziran čustven odziv, ki ga uporabnik lahko razume kot takega, čeprav to v resnici ni (Centa Strahovnik 2023b).

Posebnost spletnega sovraštva in nasilja oz. z njim povezanih čustev je tudi, da lahko presega meje spletnega in se odraža tudi v resničnem svetu. Žrtve spletnega nasilja pogosto izražajo strah in s tem povezan občutek, da je npr. nadlegovalec 'povsod', ne samo na spletu. Kar je lahko tudi res, saj medtem ko nekateri napadalci zlorabo omejujejo na omrežne tehnologije, drugi za nadlegovanje žrtev uporabljajo vsa razpoložljiva orodja, vključno s stiki v realnem prostoru. Pogosto je tudi, da čeprav se žrtve s spleta popolnoma umaknejo, čustva, povezana z nasiljem in sovraštvom, ki so ga doživele na spletu, ne minejo oz. žrtev spremljajo. Primer takega čustva je sram, povezan s 'slabim slovesom' (npr. da ima žrtev spolno prenosljivo bolezen, kar je zgolj laž nadlegovalca). Nekateri avtorji (Keats Citron 2014, 13) zato v povezavi s spletnim nadlegovanjem in zalezovanjem govorijo o sovraštvu 3.0, saj gre za personalizirano sovraštvo, ki je tako škodljivo, kot je ta nova stopnja svetovnega spleta lahko uporabna in učinkovita. Več takšnih oblik nasilja in sovraštva doživljajo ženske (zlasti gre tukaj za oblike spolnega nadlegovanja) Moški takšne napade doživijo predvsem zaradi svojih idej in dejanj. Spletni prostori ustvarjajo tudi spletne skupine, ki se lahko pretvorijo v obliko drhali. Izpostaviti velja uničujoč potencial spletnih skupin, pa tudi dinamiko sramotenja, ki je v središču nasilja in zlorab. Filozofinja Martha Nussbaum npr. izpostavlja, da so množice iz dominantnih skupin znane po tem, da sramotijo relativno nemočne skupine, da uživajo v nelagodju izključenih in stigmatiziranih (2004, 219–220). Kibernetske tolpe oz. spletna drhal se zbirajo na spletu, da bi posameznike oz. skupine nadlegovale na posebej ponižujoče in grozeče načine (Keats Citron 2014, 5).

3. Preseganje spletne sovražnosti s pomočjo sočutja in vzgoje

Z razvojem in širjenjem rabe UI so se naši načini delovanja in komuniciranja bistveno spremenili (Misnawati idr. 2025). Sklepamo lahko, da so se v digitalnem svetu posledično spremenili tudi načini in intenzivnost sovražnosti in nasilja. Čeprav zakonodajni okvir, kot sta Akt o UI, ki ga je sprejela EU (Uredba (EU) 2024/1689), in nemški Zakon o urejanju spletnih omrežij (NetzDG 2017), pomenijo pomembne

korake k regulaciji spletnih okolij in rabe UI, pa zakonodaja sama po sebi za obravnavo pojavov, pa tudi kompleksnosti škode še vedno ne zadostuje. Ker so spletna okoljih večinoma nadnacionalna, je ob hitrosti in obsegu širjenja škodljivih vsebin, nepreglednosti številnih algoritemskih sistemov to področje le z zakonodajo – ki za tehnološkimi inovacijami pogosto zaostaja – skoraj nemogoče urejati. Zato je potreben bolj celosten pristop, ki poleg pravnih vidikov in zasledovanja pravne odgovornosti terja ustrezno kultiviranje posameznikov in odpornost skupnosti. Ključno vlogo ima tu vzgoja oz. izobraževanje, ki bo v nadaljevanju tudi naš osrednji predmet zanimanja.

Naj pred obravnavo tega ožjega vidika izpostavimo, da UI lahko škoduje oz. sovražnost in nasilje še okrepi, a po drugi strani lahko predstavlja tudi del rešitve (Wang 2023; Bilewicz idr. 2021). Po eni strani lahko algoritemski sistemi sovražni govor okrepijo – toda po drugi strani nastajajoče tehnologije, kot so razločljiva UI (XAI) in orodja za moderiranje vsebin na podlagi UI (vključno z Googlovim vmesnikom API Perspective, samodejnimi detektorji, ki prepoznavajo usklajeno neavtentično vedenje, ter klasifikatorji sovražnega govora in pozivov k nasilju), kažejo, kako je sisteme UI mogoče uporabiti za prepoznavanje, preprečevanje ali blažitev digitalne sovražnosti in nasilja. Ne gre torej za zavračanje tehnologije same, ampak za njeno odgovorno rabo.

Eden izmed ključnih korakov k temu predstavlja vzgoja; v prispevku se omejujemo na vidik sočutja (v najširšem smislu) kot gradnika, ki nam pomaga omejiti oz. zaježiti in preseči pojave spletne sovražnosti in nasilja. Spletno sovraštvo se običajno razrašča v okoljih, kjer vlada anonimnost; sočutje (na primer pozivi, ki uporabnike spodbujajo, naj o vplivu svojih besed razmislijo) pa lahko sovražne interakcije ublaži. Sočutje lahko zmanjša agresivne težnje in poveča prosocialno vedenje na spletu. Sočutje spodbuja empatijo, razumevanje in vzajemno priznavanje skupne človeškosti, kar zmanjša razčlovečenje, prepreči oz. omili nekatere psihološke in družbene dejavnike spletnega sovraštva, nasilja in sovražnosti (Bilewicz idr. 2021, 261). Pri tem se ne bomo omejevali zgolj na sočutje kot čustvo oz. držo posameznika, ampak tudi na sočutje kot moralno pomembno čustvo, ki ima lahko temeljno vlogo pri doseganju pravičnosti, spodbujanju solidarnosti in oblikovanju javnih politik (Nussbaum 2013). Sočutje tako ni zgolj zasebno čustvo, temveč skupnostna drža, bistvena za etično in politično življenje – vključno z digitalno kulturo.

Pred nadaljevanjem izpostavimo še tale vidik. Stališče, da internet ali digitalni mediji predstavljajo nekakšno 'kulturo strahu' (povezano s propagando, nadzorom, širjenjem ekstremističnih vsebin, digitalnimi in razsrediščenimi mediji), je provokativno, a kot premisa v filozofskih, socioloških in medijsko-teoretskih diskurzih vse bolj aktualno. Medtem ko svetovni splet in digitalno kulturo pogosto izpostavljajo pozitivno kot demokratizirajoč prostor izmenjave informacij, kritični pogledi kažejo, da spodbuja tudi čustvene odzive – in gradi okolja, za katera so značilni tesnoba, paranoja, nadzor in moralna panika. Vse to lahko razumemo kot deviantne oblike strahu, ki so še toliko bolj lahko podlaga za vzpostavitev svojstvene kulture strahu (Strahovnik 2013). In znova – sočutje je eden izmed mehanizmov za preseganje takšnega strahu.

Če je cilj razvoj sočutja v družbi, potem mora biti to vključeno tudi v same izobraževalne poti oz. strukture. Predlog gre torej v smeri, da kot osrednji del prizadevanj za preseganje digitalnega nasilja in sovraštva uvedemo razvoj sočutja in njegovo razumevanje kot del učnega procesa. Sočutje je pomemben čustveni gradnik teh procesov, čustva pa so nekaj, kar je povezano z (ustreznimi) vrednostnimi sodbami. Digitalni svet nam velikokrat onemogoča bolj celosten vpogled v predmet lastnih čustev, vendar v znatni meri krni zmožnost ustrezne vrednostne presoje (Caspi idr. 2023). Še posebej lahko opozorimo na hitro tvorjenje čustvenih naravnosti, npr. strahu ali sovraštva, ki pogosto terja ali izzove nadaljnji hiter čustveni odziv. Zato je še bolj pomembno poudarjati pomen celostnega in izkustvenega učenja (Centa Strahovnik 2025, 138). Čustva so za vzgojo ključnega pomena, saj so neizogibni del procesa učenja in osebne rasti (Centa Strahovnik 2023). Metoda izkustvenega učenja združuje tako kognitivne kot konativne vidike in ustvarja temelje, na katerih se gojijo moralne in intelektualne vrline ter odpornost – vključno z odpornostjo v okviru digitalne kulture.

Izpostavimo lahko pet razlogov, zakaj je izkustveno učenje za digitalni svet še posebej pomembno (Centa Strahovnik 2025). Prvi razlog je, da je ta običajno izkustveno osiromašen in ne ponuja možnosti za širok razpon in pestrost izkušenj – posebej ne glede vidikov telesa in telesnosti. Tudi v primeru virtualne resničnosti takšne izkušnje ne morejo dosegati ravni izkušenj, pridobljenih v dejanskem svetu. Drugi razlog je pomanjkanje pristnih družbenih (in družabnih) vezi ali omejeno vzpostavljane takšnih vezi. Posledično je tudi manj prijateljstva, sočutja in spodbud, ki izhajajo iz medsebojne podpore pri doseganju skupnih ciljev. Pri izobraževanju za digitalni svet in o njem se zlahka osredotočimo na tehnološke vidike, medtem ko socialni vidiki ostanejo preveč v ozadju. Ena od prednosti izkustvenega učenja je, da učencem omogoča širši pogled na svet in spoštovanje skupnosti.

Tretji razlog je povezan s celostno dejavnostjo. Izkustveno učenje učence spodbuja, da sami stopijo v učno izkušnjo in tako gradijo lastno zmožnost delovanja. V spletnem okolju zlahka spregledamo pomembnost in pomen lastnega delovanja – skupaj z odgovornostjo, ki jo to prinaša. Poleg tega izkustveno učenje spodbuja vpogled v lastne sposobnosti, interese, strasti in vrednote.

Četrty, za digitalno okolje je značilno veliko motečih dejavnikov; posameznik se zlahka izgubi ali pa prednosti učenja v digitalnem svetu preprosto ne izkoristi. Izkustveno učenje medtem spodbuja motivacijo, izkušnje same pa pozornost običajno usmerjajo na določen izziv ali vprašanje.

Peti, tudi čustva in čustvene naravnosti imajo pomemben telesni gradnik, saj jih lahko pogosto zaznamo ali prepoznamo v telesu – to pa je pri celostnem in izkustvenem učenju v ospredju (Centa Strahovnik 2022).

Nenazadnje: zlahka je pretirano poudarjati tehnične ali tehnološke razsežnosti digitalne pismenosti, saj jih je na neki način preprosto učiti oz. se jih naučiti. Toda digitalni svet presega tehnične vidike in vključuje skupne prostore, komunikacijo ter zlasti pomembne etične pomisleke in vprašanja, ki jih je treba obravnavati (Strahovnik 2023; Strahovnik in Centa Strahovnik 2024; Žalec 2023). Učenje oz.

vzgoja se mora povezovati s trenutnimi izkušnjami, ki vključujejo tudi moralno intuicijo in sodbe ali – širše – etična stališča, ki oblikujejo naše življenje. Tako lahko svojo trenutno izkušnjo ali položaj in življenje kot celoto 'odigramo', povežemo in tudi razumemo (Allison idr. 2017; Centa 2018).

4. Zaključek

Sočutje je čustvo, ki ga lahko razumemo kot tisto, ki širi meje lastnega jaza. Je čustvo, ki nas povezuje z drugim, saj oseba z drugim sočustvuje le, kadar npr. trpljenje drugega lahko dojamemo kot pomemben del svoje sheme ciljev in projektov. To pomeni, da trpljenje drugega dojamemo kot napad na lasten razcvet oz. dobro. Nadalje, oseba, ki sočustvuje, mora v drugi »trpeči« osebi (u)videti lastno ranljivost (Nussbaum 2001).

V digitalni kulturi je gojiti in vzgajati občutje za sočloveka še toliko bolj pomembno, saj že v resničnem življenju želja pomagati drugemu in izkušnja sočutja pogosto nista dovolj vztrajni in močni, da bi prešli v ustrezna dejanja. Zato Martha Nussbaum poudarja praktični vidik sočutja, kar pa pomeni, da moramo tudi skupnost usmerjati in oblikovati tako, da se bomo na podlagi sočutja lahko ustrezno odzivali. Odnos med posamezniki in družbenimi ali skupnostnimi ustanovami je tu dvosmeren. Sočutni posamezniki sestavljajo določeno skupnost, ki uteleša to, kar si ti posamezniki zamišljajo, obenem pa kot skupnost povratno vplivajo na razvoj sočutja in ustreznega ravnanja še pri drugih posameznikih (2001, 405). Zato v okviru svojih predlogov Nussbaum poudarja pomen vzgoje za sočutje. Ta mora vključevati krepitev sposobnosti življenja v drugega. Poudarek je na posamezniku, ki na svet okoli sebe, na tok zgodovine, na ljudi, ki ga obkrožajo, gleda s sočutjem. Takšne vzgojne vsebine pri posamezniku vzbujajo sočutno zamišljanje in primerjanje svojega razumevanja dobrega z razumevanji drugih.

Krščanska misel in krščanska teologija o sočutju v digitalni kulturi izpostavljata, da je ob zavedanju o edinstvenosti človeka (ki je UI zaenkrat ne more poustvariti) pomembna tudi sočutna skupnost, ki temelji na pristnih odnosih, izhajajočih iz sočutne vzgoje. Ta skupnost je prostor, v katerem sta moralna in duhovna identiteta ključni izhodišči za vzgojo ozaveščenega, moralno odgovornega posameznika – vključno z njegovo umeščenostjo v digitalno kulturo. Okvir krščanske teologije ponuja večplastno razumevanje sočutja, ki temelji na Božjem razodetju o usmiljenem, milostnem Bogu (2 Mz 34,5-7; Iz 63,7-9), vzoru in vzorcih Božje solidarnosti oz. sočutja s trpljenjem človeka (Mt 9,35-38) in praksah, ki poudarjajo dejanja usmiljenja. Sočutje je hkrati Božje določilo in dolžnost človeka. Na podlagi razumevanja sočutja kot oblike ljubezni in kot vrline, ki dopolnjuje dobrotnost, nekateri sodobnejši misleci te tradicije to razumevanje nadgrajujejo v smeri, da sočutje povezujejo tudi s solidarnostjo z zatiranimi, nenasiljem, pravičnostjo in pa osebno pozornostjo do trpečih (prim. Hauerwas 1983; Francišek 2020; Žalec 2022). Teološka refleksija digitalne kulture lahko to plodno uporabi v okoljih, ki jih zaznamujejo polarizacija, nadlegovanje in algoritmično ojačano ogorčenje, ter predlaga prakse resničnega,

zavestnega in obnovitvenega sočutja. Odziv oz. odzivnost, ki jo vodi sočutje, sledi načelu, da drugega najprej ‚pogledamo‘, preden odgovorimo oz. se odzovemo – upoštevajmo njegovo ali njeno zgodbo, zlasti pa ranljivost. To ne pomeni odpovedi govoriti po in o resnici, nasprotno. Kakor je zapisano v pismu Efežanom (4,25-32):

»Zato opustite laž in govorite resnico vsak s svojim bližnjim, saj smo med seboj deli enega telesa. *Jezite se, a nikar ne grešite*; sonce naj ne zaide nad vašo jezo in ne dajajte prostora hudiču. /.../ Nobena umazana beseda naj ne pride iz vaših ust, marveč le dobra, da bi bila ob potrebi v izgrajevanje, da bi podelila milost tistim, ki poslušajo. Ne žalostite Božjega Svetega Duha, s katerim ste bili kot s pečatom zaznamovani za dan odkupitve. Naj izginejo med vami vsakršna ujedljivost, vsakršno besnenje, jeza, rohnenje in preklinjanje z vsakršno hudobijo vred. Bodite drug do drugega dobrosrčni in usmiljeni ter drug drugemu odpuščajte, kakor je tudi vam Bog milostno odpustil v Kristusu.«

Hkrati je v tem razumevanju sočutja izpostavljena pomembnost moralne bližine – za razliko od fizične oddaljenosti kot določila našega delovanja. To je še posebej pomembno v digitalni kulturi, kjer se zlahka ustvarjajo prostori fizične in družbene oddaljenosti, za katero se zdi, da posamezniku daje lažen občutek moralne sprejemljivosti spletnega sovraštva in nasilja (Bauman 1995; Vetlesen 1994; Volf 1996). Že Bauman je opozarjal na spremembe dojemanja nasilja v kontekstu, kadar „žrtvam ni treba gledati v oči; ko preštevamo zgolj pike na zaslonih, ne pa trupel.“ (1995, 150)

Znova: ne gre za to, da bi na podlagi teh pomislekov in s pozivi k celostnemu, izkustvenemu učenju pozivali k zavračanju ali izključevanju tehnologije – vključno z digitalnimi tehnologijami. Deloma lahko namreč pri poučevanju sočutja pomaga tudi premišljena in skrbno načrtovana uvedba novih tehnologij v izobraževanje. Jeglič tako ugotavlja, da bi bilo s „pomočjo virtualne resničnosti (VR) /.../ smiselno poučevati tudi o religijskih in etičnih vsebinah. Uporaba VR lahko posamezniku pomaga, da se vživi v situacijo drugega, za kar pa mora biti seveda tudi dovolj empatičen in nenazadnje sočuten.“ (Jeglič 2024, 926–7; prim. Globokar 2024)

Pri prepoznavi, preprečevanju in omejevanju posledic spletnega nasilja in sovravnosti je torej bistveno, da vključimo več premislekov. Najprej se moramo vprašati, kdo je ‚drugi‘ v digitalnem okolju, saj razumevanje sočloveka oblikuje temelje za odgovorno komunikacijo in preostale vidike delovanja. Nato se je treba ustaviti ob vprašanju, kako in kaj v digitalnem prostoru pomeni nekoga resnično ‚videti‘ in ‚slišati‘, saj to presega zaznavanje vsebin in zahteva priznavanje dostojanstva ter občutljivosti sogovornika. Nenazadnje pa je nujno ozaveščanje in prevzemanje odgovornosti za lastno delovanje, ki vključuje tudi spletna in digitalna dejanja.

Reference

- Allison, Pete, in Jayson Seaman.** 2017. *Experiential Education*. V: Michael Peters, ur. *Encyclopedia of Educational Philosophy and Theory*. Singapore: Springer Nature. https://doi.org/10.1007/978-981-287-532-7_449-1
- Adamopoulou, Eleni, in Lefteris Moussiades.** 2020. An Overview of Chatbot Technology. V: Ilias Maglogiannis, Lazaros Iliadis in Elias Pimenidis, ur. *IFIP Advances in Information and Communication Technology*, 584. Cham: Springer. https://doi.org/10.1007/978-3-030-49186-4_31
- Abidin, Crystal.** 2020. Meme Factory Cultures and Content Pivoting in Singapore and Malaysia during COVID-19. *Misinformation Review* 1. <https://doi.org/10.37016/mr-2020-031>
- Aceto, Giuseppe, in Antonio Pescapé.** 2015. Internet Censorship Detection: A Survey. *Computer Networks* 83:381–421. <https://doi.org/10.1016/j.comnet.2015.03.008>
- Aiken, Mary.** 2016. *The Cyber Effect: A Pioneering Cyberpsychologist Explains How Human Behaviour Changes Online*. New York: Spiegel & Grau.
- Aly, Anne, Stuart Macdonald, Lee Jarvis in Thomas Chen.** 2016. *Violent Extremism Online: New Perspectives on Terrorism and the Internet*. London: Routledge.
- Bauman, Zygmund.** 1995. *Life in Fragments: Essays in Postmodern Morality*. Oxford: Blackwell.
- Bermudez-Villalva, Adrian, Maryam Mehrnezhad in Ehsan Toreini.** 2025. Measuring Online Hate on 4chan Using Pre-Trained Deep Learning Models. *IEEE Transactions on Technology and Society* 6, št. 2:200–209. <https://doi.org/10.1109/TTS.2025.3549931>
- Bilewicz, Michał, Tempeska Patrycja, Leliwa Gniewosz, Maria Dowgiałło, Michalina Tańska, Rafał Urbaniak in Michał Wroczyński.** 2021. Artificial Intelligence against Hate: Intervention reducing Verbal Aggression in the Social Network Environment. *Aggressive Behavior* 47:260–266. <https://doi.org/10.1002/ab.21948>
- Brady, William J., Killian L. McLoughlin, Mark P. Torres, Kara F. Luo, Maria Gendron in Molly J. Crockett.** 2023. Overperception of moral Outrage in Online Social Networks inflates Beliefs about Intergroup Hostility. *Nature Human Behaviour* 7:917–927. <https://doi.org/10.1038/s41562-023-01582-0>
- Caspi, Avner, in Shir Etgar.** 2023. Exaggeration of Emotional Responses in Online Communication. *Computers in Human Behavior* 146. <https://doi.org/10.1016/j.chb.2023.107818>
- Centa, Mateja.** 2018. Kognitivna teorija čustev, vrednostne sodbe in moralnost. *Bogoslovni vestnik* 78, št. 1:53–65.
- Centa Strahovnik, Mateja.** 2022. Spoznavna vloga telesnih občutkov, resonanca in odtujenost *Bogoslovni vestnik* 82, št. 3:611–621. <http://doi.org/10.34291/BV2022/03/Centa>
- . 2023. Čustva, človekova odnosnost in doseganje dobrega življenja. Ljubljana: Teološka fakulteta.
- . 2023b. Identiteta in pogovorni sistemi umetne inteligence. *Bogoslovni vestnik* 83, št. 4:853–864. <https://doi.org/10.34291/BV2023/04/Centa>
- . 2025. Sovražstvo in sočutje v digitalnem svetu: identiteta, umetna inteligenca in vrline. V: *Sočutni jezik med mladimi*, 133–140. Ljubljana: Urad RS za mladino, Ministrstvo za vzgojo in izobraževanje,
- Fire, Michael, Yitzhak Elbazi, Adi Wasenstein in Lior Rokach.** 2025. Dark LLMs: The Growing Threat of Unaligned AI Models. *Arxiv*. <https://doi.org/10.48550/arXiv.2505.10066>
- Frančiček.** 2020. Fratelli tutti. Vsi smo bratje. Okrožnica. Vatican, 3. 10. https://www.vatican.va/content/francesco/sl/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html
- Globokar, Roman.** 2024. Pogled Katoliške Cerkve na razvoj in uporabo umetne inteligence. *Bogoslovni vestnik* 84, št. 4:867–883.
- Hauerwas, Stanley.** 1983. *The Peaceable Kingdom: A Primer in Christian Ethics*. Bloomington: University of Notre Dame Press.
- Hietanen, Mika, in Johan Eddebo.** 2022. Towards a Definition of Hate Speech—With a Focus on Online Contexts. *Journal of Communication Inquiry* 47, št. 4:440–458. <https://doi.org/10.1177/01968599221124309>
- Jeglič, Urška.** 2024. Virtualna resničnost kot pripomoček pri poučevanju religijskih vsebin v šolah. *Bogoslovni vestnik* 84, št. 4:921–930. <https://doi.org/10.34291/BV2024/04/Jeglič>
- Keats Citron, Danielle.** 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.
- . 2015. Addressing Cyber Harassment: An Overview of Hate Crimes in Cyberspace. *Journal of Law, Technology & the Internet* 6, št. 1:1–12.
- Mahl, Daniela, Sabrina H. Kessler, Mike S. Schäfer, Anna Jobin, Jing Zeng in Fanny Georgi.** 2025. Conspiracy Theories and Misinformation in Digital Media: An International Expert Assessment of Challenges, Trends, and Interventions. *Communications: pred tiskom*. <https://doi.org/10.1515/commun-2024-0123>
- Maras, Marie-Helen, in Alex Alexandrou.** 2018. Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos. *The International Journal of Evidence & Proof* 23, št. 3:255–262. <https://doi.org/10.1177/13657127188072>
- Marino, Erik Bran, Jesus M. Benitez-Baleato in Ana Sofia Ribeiro.** 2024. The Polarization Loop: How Emotions Drive Propagation of Disinfor-

- mation in Online Media—The Case of Conspiracy Theories and Extreme Right Movements in Southern Europe. *Social Sciences* 13, št. 11:603. <https://doi.org/10.3390/socsci13110603>
- Misnawati, Desy, Farida Olivia Ridheta Citrawijaya in Rosalia Dewi Arlusi.** 2025. The Impact of Artificial Intelligence on Human Interaction. *The Journal of Academic Science* 2, št. 1. <https://doi.org/10.59613/gso0mmj59>
- Mostafazadeh Davani, Aida, Mohammad Atari, Brendan Kennedy in Morteza Dehghani.** 2023. Hate Speech Classifiers Learn Normative Social Stereotypes. *Transactions of the Association for Computational Linguistics* 11:300–319. https://doi.org/10.1162/tacl_a_00550
- NetzDG 2017 (Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken).** 2017. Web Archive. https://web.archive.org/web/20220630204044/https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html (pridobljeno 22. 4. 2024).
- Nussbaum, C. Martha.** 2001. *Upheavals of Thought: The Intelligence of Emotions*. New York: Cambridge University Press.
- . 2004. *Hiding from Humanity. Disgust, Shame, and the Law*. New Jersey. Princeton, NJ: Princeton University Press.
- . 2013. *Political Emotions: Why Love Matters for Justice*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Patel Harikrishna in Hussain Faiza.** 2024. Do AI Chatbots Incite Harmful Behaviours in Mental Health Patients? *BJPsych Open* 10, št. 1:70–71. <https://doi.org/10.1192/bjo.2024.225>
- Ramponi, Alan, Benedetta Testa, Sara Tonelli in Elisabetta Jezek.** 2022. Addressing religious hate online: from taxonomy creation to automated detection. *PeerJ Computer Science* 8:e1128. <http://doi.org/10.7717/peerj-cs.1128>
- Slonje, Robert in Peter K. Smith.** 2008. Cyberbullying: Another Main Type of Bullying?. *Scandinavian Journal of Psychology* 49:147–154. <https://doi.org/10.1111/j.1467-9450.2007.00611.x>
- Sarkar, Advait.** 2025. AI Could Have Written This: Birth of a Classist Slur in Knowledge Work. *CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, članek št. 621:1–12. <https://doi.org/10.1145/3706599.3716239>
- Strahovnik, Vojko.** 2013. Beyond the Culture of Fear: Fear and Responsibility. V: Janez Juhant in Bojan Žalec, ur. *From Culture of Fear to Society of Trust*, 91–98. Zürich: Lit.
- . 2023. Etični in teološki izzivi velikih jezikovnih modelov. *Bogoslovni vestnik* 83, št. 2:839–852.
- . 2024. Algoritemsko nasilje. *Bogoslovni vestnik* 84, št. 4:783–795.
- Strahovnik, Vojko, in Mateja Centa Strahovnik.** 2024. Transhumanism, Human Moral Enhancement, and Virtues. *Religions* 15, št. 11:1345. <https://doi.org/10.3390/rel15111345>
- Steffensen, Torjus, Sampsa Kohtala, Håvard Vestad in Martin Steinert.** 2020. TrollBOT: A Spontaneous Networking Tool Facilitating Rapid Prototyping of Wirelessly Communicating Products. *Procedia CIRP* 91:634–638. <https://doi.org/10.1016/j.procir.2020.03.111>
- Thomas, Kurt, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, Damon McCoy, Sarah Meiklejohn, Thomas Ristenpart in Gianluca Stringhini.** 2021. SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. V: *2021 IEEE Symposium on Security and Privacy (SP)*, 247–267. San Francisco, CA: [b. z.]. <https://doi.org/10.1109/SP40001.2021.00028>
- Thompson, L. Robin.** 2011. Radicalization and the Use of Social Media. *Journal of Strategic Security* 4, št. 4:167–190. <http://dx.doi.org/10.5038/1944-0472.4.4.8>
- Uredba (EU) 2024/1689 evropskega parlamenta in sveta o določitvi harmoniziranih pravi o umetni inteligenci in spremembi uredb (ES).** 2024. <http://data.europa.eu/eli/reg/2024/1689/oj>
- Vetlesen, Arne Johan.** 1994. *Perception, Empathy, and Judgment*. University Park: Pennsylvania State University Press.
- Volf, Miroslav.** 1996. *Exclusion and Embrace: A Theological Exploration of Identity, Otherness, and Reconciliation*. Nashville: Abingdon.
- Wang, Qizheng.** 2023. Advanced Deep Learning Approaches for Hate Speech Detection. *Highlights in Science, Engineering and Technology*, št. 41:158–164. <https://doi.org/10.54097/hset.v41i.6801>
- Westerlund, Mika.** 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review* 9, št. 11:40–53. <http://doi.org/10.22215/timreview/1282>
- Završnik, Aleš in Louise Hooper.** 2025. *Digitalna razsežnost nasilja nad ženskami in dekleti v Sloveniji*. Ljubljana: Svet Evrope.
- Žalec, Bojan.** 2022. Resonanca, demokracija in skupno dobro. *Bogoslovni vestnik* 82, št. 3:521–533.
- . 2023. *Personalistična filozofija edukacije: odnosi in etika v vzgoji in izobraževanju*. Ljubljana: Teološka fakulteta.
- Ybarra, L. Michele, Kimberly J. Mitchell, Janis Wolak in David Finkelhor.** 2006. Examining Characteristics and Associated Distress Related to Internet Harassment. *Pediatrics* 118, št. 4:1169–1177. <https://doi.org/10.1542/peds.2006-0815>