



**Developing Community Media to Mitigate  
the Impact of Fake News**

**Modul 2 -Modeli  
moderiranja i  
primjene digitalnih  
sadržaja**



**Pozdrav!**

## **Pregled**

Moderiranje sadržaja postupak je kojem se posvećuje ograničena pažnja i koji često prolazi nezapaženo u našem internetskom iskustvu. Istodobno se smatra potrebnim kako bi se korisnicima osiguralo sigurno pregledavanje i iskustvo.

Kroz aktivnosti ove radionice dobit ćete pregled pojmova te pitanja i pristupe koji oni uključuju

## **Teme**

- Što je moderiranje sadržaja i njegova važnost
- Modeli moderiranja sadržaja
- Tehnički alati i pristupi u moderiranju sadržaja
- Crne liste i filtri za objavljivanje postova
- Praksa moderiranja sadržaja na društvenim mrežama

Radionica će nastojati obuhvatiti sljedeće teme:

- Što je moderiranje sadržaja i njegova važnost
- Modeli moderiranja sadržaja
- Tehnički alati i pristupi u moderiranju sadržaja
- Crne liste i filtri za objavljivanje postova
- Praksa moderiranja sadržaja na društvenim mrežama

## **Ishodi učenja**

- Shvatiti što je moderiranja sadržaja i njegova nužnost
- Upoznati se s različitim vrstama moderiranja i njihovim prednostima
- Saznati kako stvoriti "crnu listu" i primijeniti filtre nakon objave
- Razumijeti praksu koju pružatelji usluga socijalnih medija koriste kako bi identificirali anomalije u širenju

U sklopu obuke koja se nudi i materijala za samosumjereno učenje polaznici će:

Upoznati se s različitim vrstama moderiranja i njihovim prednostima

Steći osnovno znanje o raznim tehničkim alatima

Moći prepoznati karakteristike negativnog i provokativnog sadržaja koji se namjerno objavljuje na društvenim mrežama

Saznati kako stvoriti "crnu listu" i primijeniti filtre nakon objave

Razumijeti praksu koju pružatelji usluga socijalnih medija koriste kako bi identificirali anomalije u širenju

## **Prezentacija teorije**

Što je moderiranje sadržaja i  
njegova važnost



## Uvod

- Širenje digitalnih tehnologija širenjem interneta dovelo je do 'demokratizacije informacija'
- Građani se sada za informacije ne oslanjaju samo na tradicionalne medije (TV, radio i novine)
- Imaju pristup obilju izvora, kao i sposobnost razmjene vlastitih misli i iskustava

U sklopu obuke koja se nudi i materijala za samosumjereno učenje polaznici će:

Upoznati se s različitim vrstama moderiranja i njihovim prednostima

Steći osnovno znanje o raznim tehničkim alatima

Moći prepoznati karakteristike negativnog i provokativnog sadržaja koji se namjerno objavljuje na društvenim mrežama

Saznati kako stvoriti "crnu listu" i primijeniti filtre nakon objave

Razumijeti praksu koju pružatelji usluga socijalnih medija koriste kako bi identificirali anomalije u širenju

## Uvod

- Unatoč početnom optimizmu, ovaj se razvoj više ne smatra potpuno pozitivnim
- Internet je postao prostor u kojem se dijele prijeteći i zlonamjerni materijali.
- Potrebno je poduzeti korake kako bi se osiguralo da sadržaj koji korisnici dijele i kojem pristupaju nije štetan.

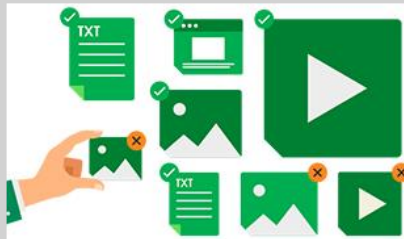
Dočekali smo ovaj razvoj događaja s velikim uzbuđenjem i nadom. Mnogi su analitičari ovu evoluciju doživjeli kao svitanje velikog doba u kojem će pojedinci biti više informirani, a javni će diskurs eksponencijalno profitirati aktivnim sudjelovanjem svih građana i zastupanjem svih gledišta. Međutim, ovaj je optimizam ubrzo slomljen jer je digitalna sfera također ponudila plodno tlo za razmjenu i procvat prijetećeg i često mržnjom ispunjenog materijala, kao i za širenje neprovjerenih i često namjernih lažnih informacija.

Slijedom toga, od iznimne je važnosti da svaki oblik medija uvede i uspostavi postupke moderiranja sadržaja kako bi se osiguralo da sadržaj koji se izlaže nije štetan. Potrebni su konkretni koraci kako bi se osiguralo blokiranje potencijalno štetnog sadržaja u obliku, na primjer, materijala za zlostavljanje djece, nasilnog i ekstremnog sadržaja, govora mržnje, grafičkog sadržaja, seksualnog sadržaja, okrutnog i neosjetljivog materijala i neželjenog sadržaja (Ofcom: 2019). U tu svrhu, ovaj modul želi predstaviti specifične modele za moderiranje medija s resursima kako bi podržao polaznike koji su zainteresirani za razvoj medija u zajednici da učinkovito provode takve procese u svom radu.



## Prezentacija teorije

## Stavljanje moderiranja sadržaja u perspektivu



## Što je moderiranje sadržaja?

- Rječnik Cambridge definira moderiranje kao kvalitetu izvođenja nečeg unutar razumnih granica.
- U kontekstu medija ovo se načelo odnosi na postavljanje propisa i granica u pogledu informacija i materijala koji će se dijeliti ili hostirati
- Odnosi se na nadgledanje sadržaja koji se prenosi i blokiranje ili uklanjanje bilo kojeg sadržaja koji se smatra neprihvatljivim

Cambridge Dictionary definira moderiranje kao kvalitetu izvođenja nečeg unutar razumnih granica. U kontekstu medijskog sadržaja, ovo se načelo odnosi na postavljanje propisa i ograničenja informacija i materijala koji će se dijeliti na platformi koja je medij, internetska stranica društvenih mreža, internetska stranica ili blog i poduzimati korake u kako bi se osiguralo poštivanje istih. Drugim riječima, odnosi se na nadgledanje sadržaja koji je već postavljen ili će tek biti postavljen i ne dopustiti ili ukloniti bilo koji sadržaj koji nije prihvatljiv prema skupu pravila po kojima dotična platforma postoji (Grimes-Viort: 2010). „Ovisno o specifičnim zahtjevima svake internetske stranice, moderiranje se može odviti u većoj ili manjoj mjeri. Neki se internetski forumi ponose slobodom govora, dok druge, poput društvenih mreža, moraju uspostaviti tešku ravnotežu između jednostavnosti upotrebe i zaštite svojih mlađih korisnika“ (Smith: 2019).

## Što je moderiranje sadržaja?

- Velika većina platformi i medija (ili digitalnih ili tradicionalnih) na društvenim mrežama imaju posebne smjernice za moderiranje koje u određenoj mjeri diktiraju koja je vrsta sadržaja prihvatljiva za širenje
- Poštujući ova pravila i primjenjujući različite tehnike i pristupe moderiranja sadržaja, oni nadgledaju i ocjenjuju materijal podijeljen s njima ili putem njih

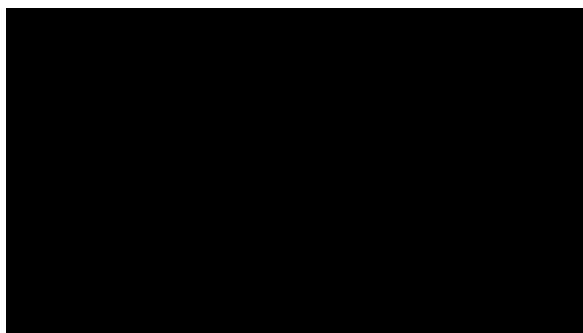
Većina popularnih društvenih medija i medijskih kuća (bilo digitalnih ili tradicionalnih) imaju posebne smjernice za moderiranje koje u određenoj mjeri diktiraju koja je vrsta sadržaja prihvatljiva za širenje putem njihove platforme. Pridržavajući se ovih pravila i primjenjujući različite tehnike i pristupe moderiranju sadržaja, oni nadgledaju i procjenjuju dijeljeni materijal kako bi osigurali da se ne objavljuju ili se uklanjaju sve vrste uvredljivih ili neprikladnih sadržaja poput porno videa, eksplicitnih slika ili slika koje nisu prikladne za sve dobne skupine ljudi. (Cogito Tech: 2020).

### **Zašto je potrebno moderiranje sadržaja?**

- Ključno je osigurati da publika / korisnici medija ili platforme ne budu izloženi bilo čemu neprikladnom ili štetnom
- Hostiranje ili dijeljenje bilo kakvog neprimjerenog, štetnog ili ilegalnog ponašanja ne samo da će stvarati probleme platformi / medijima, već može imati ozbiljne implikacije na njihovu publiku, posebno na mlađu dob
- Stoga je vrlo važno da dotični mediji i platforme budu na oprezu

Moderiranje slika, video zapisa i teksta presudni su kako bi se osiguralo da publika / korisnici medija ili platforme ne budu izloženi bilo čemu neprikladnom ili štetnom (Schomer: 2019). Dijeljenje bilo kakvog neprimjerenog, štetnog ili nezakonitog ponašanja ne samo da će stvarati probleme platformi / medijima, već može imati ozbiljne implikacije na njihovu publiku, posebno onu mlađe dobi koja koristi internet i pretežno društvene medije za pristup informacijama i vijestima kao i za socijalne interakcije. Stoga je od najveće važnosti da su ti mediji i platforme oprezni i ulažu u praćenje i moderiranje sadržaja koji je dostupan putem njih (puremoderation: 2020).

## **Stavljanje moderiranja sadržaja u perspektivu**





**AKTIVNOST Br. I**  
**Prijetnje koje moderatori moraju riješiti**

Putem svojih telefona pošaljite na mentimeter prijetnje kojih se možete sjetiti

## **Teorija - Presentacija: Modeli moderacije sadržaja**



Postoje različiti pristupi ili modeli moderiranja sadržaja koje platforma može slijediti

Ovi se modeli u određenoj mjeri međusobno ne isključuju pa se može odlučiti za uspostavljanje više pristupa.

U konačnici, mora odlučiti koji je najučinkovitiji i za to najbolje funkcionira

Kada je riječ o moderiranju sadržaja, postoje različiti pristupi ili modeli koje platforma može upotrijebiti ovisno o vlastitim načelima, načinu rada, ciljevima i radnim procesima. Ovi se modeli u određenoj mjeri međusobno ne isključuju pa se može odlučiti za uspostavljanje više pristupa. U konačnici, za pristup se treba odlučiti s ciljem osiguravanja najboljeg mogućeg ishoda praćenja i filtriranja koji će osigurati njegovu publiku od štetnih i nezadovoljnih sadržaja, kao i zadržati određeni osjećaj reda u svom radu (Cogito Tech: 2019).

Razumijevanje različitih vrsta moderiranja sadržaja, zajedno s njihovim snagama i slabostima, može pomoći moderatoru medija da donese pravu odluku koja će najbolje odgovarati njegovoj svrsi i načinu rada.





||UNTRANSLATED\_CONTENT\_START||As the name signifies, the Pre-moderation model involves setting a process in place where all the content requires approval before it is published. ||UNTRANSLATED\_CONTENT\_END||U praktičnom smislu, sav sadržaj koji se pošalje platformi stavlja se na čekanje da ga moderator pregleda i odobri prije nego što postane vidljiv publici korisnika. Pred-moderiranje pomaže osigurati da se (u rukama dobrog moderatora) bilo koji sadržaj koji može biti štetan, neprikladan ili ne poštuje pravila internetske stranice / platforme drži izvan vidljivih javnih odjeljaka. Stoga nudi visoku kontrolu upravitelju-moderatoru.

## **Nedostaci pred-moderacije**

- **To može uzrokovati pritužbe pojedinca koji podnosi sadržaj**
- To može dovesti do kašnjenja koja mogu negativno utjecati na sadržaj koji je konverzacijski i vremenski osjetljiv
- To podrazumijeva veliki napor kojim može postati teško upravljati

Ipak ima svojih loših strana jer može uzrokovati pritužbe pojedinca koji podnosi sadržaj ili može dovesti do kašnjenja u prijenosu što može imati negativan učinak na sadržaj koji je konverzacijski i vremenski osjetljiv. |||UNTRANSLATED\_CONTENT\_START||| Additionally, this is a process that entails a high cost if and when a platform grows and submissions cross a threshold of user-generated content unmanageable by moderators (Grimes-Viort: 2010). |||UNTRANSLATED\_CONTENT\_END|||

## Post-moderiranje

- Omogućava objavljivanje sadržaja / materijala na platformi bez ikakvog početnog pregleda, a zatim ga procijenite
- Minimizira kašnjenja i pomaže bržem radu platforme i angažmana koji se odvijaju na / kroz nju
- Omogućuje moderatorima da osiguraju sigurnost, dok se problemi u ponašanju i zakoni mogu pravovremeno identificirati i na njih se može postupiti



Izvor slike: [www.telusinternational.com/articles/mastering-content-moderation](http://www.telusinternational.com/articles/mastering-content-moderation)

Ovaj pristup uključuje omogućavanje objavljivanja sadržaja / materijala na platformi bez ikakvog početnog pregleda, a moderatori ga mogu procijeniti i po potrebi ukloniti. Prednost ove vrste moderiranja je što smanjuje kašnjenja i pomaže bržem radu platforme i angažmana koji se odvijaju na / kroz nju. "Ljudi očekuju razinu neposrednosti prilikom interakcije na internetu, a moderiranje objava to omogućava, a istovremeno moderatorima omogućuje da osiguraju da se sigurnosni, bihevioralni i pravni problemi mogu prepoznati i pravovremeno postupati prema njima" (Grimes-Viort: 2010.).

### **Nedostaci post-moderiranja**

- Ipak, to može dovesti do toga da neki korisnici naiđu na štetan ili neprimjeren sadržaj.
- Štoviše, to je proces koji može postati skuplji kako raste vidljivost i doseg platforme, a kao rezultat povećava se sadržaj koji joj se podnosi.

Ipak, to može dovesti do toga da neki korisnici naiđu na štetan ili neprimjeren sadržaj. Štoviše, to je proces koji može postati skuplji kako raste vidljivost i doseg platforme, a kao rezultat povećava se sadržaj koji joj se podnosi.



**Reaktivno moderiranje**

- Reaktivno moderiranje pristup je koji se oslanja na korisnike da prijave sadržaj kad ga smatraju neprikladnim.
- Uspostavljen je postupak izvještavanja koji korisnici mogu označiti sadržaj i zatražiti njegovo uklanjanje.
- Može se opisati kao moderiranje sadržaja koji generiraju korisnici i nedvojbeno daje moć i vlasništvo u rukama korisnika-publike platforme (Cogito Tech: 2019).

Izvor slike: [www.the-future-of-commerce.com/2019/10/17/social-media-and-customer-service-alignment/](http://www.the-future-of-commerce.com/2019/10/17/social-media-and-customer-service-alignment/)

Reaktivno moderiranje pristup je koji se oslanja na korisnike da prijave sadržaj kad ga smatraju neprikladnim. Uspostavljen je postupak izvještavanja koji korisnicima pruža mogućnost obavijestavanja administratora ili moderatora i traženja da se navedeni sadržaj pregleda i ukloni. Reaktivno moderiranje široko se koristi u društvenim mrežama gdje korisnici mogu aktivno pridonijeti moderiranju sadržaja. Može se opisati kao moderiranje sadržaja koji generiraju korisnici i nedvojbeno daje moć i vlasništvo u rukama korisnika-publike platforme (Cogito Tech: 2019).

## **Nedostaci reaktivne umjerenosti**

- To može dovesti do pretjerane ovisnosti o korisnicima i ograničene kontrole.
- Moderiranje sadržaja postaje pitanje interpretacije i rasprave između korisnika.

Glavna prednost ove metode moderiranja je u tome što se može prilagoditi vidljivosti platforme i omogućiti joj da izbjegne odgovornost za neprikladan sadržaj koji su korisnici prenijeli sve dok odgovara na zahtjeve za preispitivanje i uklanjanje takvog materijala. Međutim, to povlači rizik jer dovodi do pretjerane ovisnosti o korisnicima (Grimes-Viort: 2010.).

## Distribuirano moderiranje

- Korisnički način moderiranja sadržaja koji se oslanja na sustav ocjenjivanja
- Korisnici glasaju o tome jesu li prijave u skladu s očekivanjima zajednice ili u skladu s pravilima korištenja
- Na taj način doprinosi, uz podršku angažiranih i iskusnih moderatora, u kontroli komentara ili postova na forumima.

A vibrant, abstract illustration representing the challenges of moderating user-generated content. It features a central laptop with two hands typing on the keyboard. Surrounding the laptop are various social media and digital communication icons, including a smartphone, a speech bubble, a power button, a location pin, a smiley face, an '@' symbol, a mail icon, a camera, a play button, a heart, a star, a share icon, and a thumbs up. The background is a mix of blue and purple hues with a grid-like pattern. The text 'USER GENERATED CONTENT' is prominently displayed in the upper right, and 'STOP' is written vertically on the left. The word 'LIKE' is also visible near the laptop. The overall theme is digital interaction and community moderation.

Izvor slike: [www.eunazi.com/user-generated-content-moderation-challenges-beyond/](http://www.eunazi.com/user-generated-content-moderation-challenges-beyond/)

Korisnička metoda moderiranja sadržaja oslanjajući se na sustav ocjenjivanja koji članovi zajednice koriste da glasaju jesu li prijave u skladu s očekivanjima zajednice ili u skladu s pravilima korištenja. Na taj način doprinosi, uz podršku angažiranih i iskusnih moderatora, u kontroli komentara ili postova na forumima.

## **Nedostaci distribuiranog sadržaja**

- Previše se oslanja na sposobnost i spremnost internetske zajednice da se samostalno moderira

Važno je uzeti u obzir da se ovaj pristup oslanja na sposobnost i spremnost internetske zajednice da se samomoderira.



### **Automatizirano moderiranje**

- Uključuje uvođenje i upotrebu tehničkih alata i automatiziranih procesa filtriranja za provjeru i pregled sadržaja i prijava.
- Najčešći alat koji se koristi je filter riječi koji ispituje tekst za prethodno definirane zabranjene riječi i zamjenjuje ih ili blokira cijeli tekst.
- Slični alati postoje za moderiranje slika i videozapisa koji označavaju neprikladan sadržaj navedenih multimedijских sadržaja.

Uz sve gore navedene sustave za moderiranje koje pokreću ljudi, automatizirano moderiranje dragocjeno je oružje u arsenalu moderatora. Uključuje uvođenje i upotrebu tehničkih alata i automatiziranih procesa filtriranja za provjeru i pregled sadržaja i prijava. Najčešći alat koji se koristi je filter riječi koji ispituje tekst za prethodno definirane zabranjene riječi i zamjenjuje ih ili blokira cijeli tekst. Slični alati postoje za moderiranje slika i videozapisa koji označavaju neprikladan sadržaj navedenih multimedijских sadržaja. (Grimes-Viort: 2010.).

## **Nedostaci automatiziranog moderiranja**

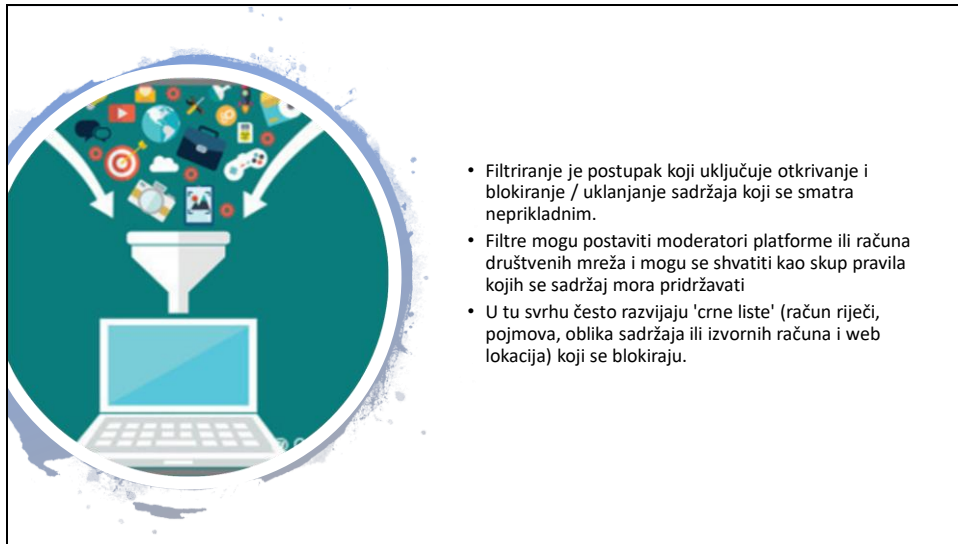
- Automatizirani sustavi mogu se prevariti ili neće uspjeti ukloniti automatski štetni ili prijeteći materijal
- Možda će im nedostajati kontekst ili pristup materijalu koji pregledavaju



AKTIVNOST Br. 2

Stvaranje crne liste i postavljanje filtera  
publikacije

Putem svojih telefona pošaljite na mentimeter prijetnje kojih se možete sjetiti



Filtriranje je postupak koji uključuje otkrivanje i blokiranje / uklanjanje sadržaja koji se smatra neprikladnim. Filtre mogu postaviti moderatori platforme ili profila društvenih mreža, a to je skup pravila kojih se sadržaj mora pridržavati ako će biti objavljen ili ostati vidljiv putem platforme (Vijeće Europe: 2017). U tu svrhu često razvijaju 'crne liste' (riječi, pojmovi, oblici sadržaja ili izvornih profila i internetskih stranica) kojima je blokirano objavljivanje, dijeljenje ili komentiranje na platformi ili računu.

- Drugim riječima, moderatori sadržaja mogu sastaviti arhivu riječi ili izraza koji nisu dopušteni za njihovu platformu, kao i popis internetskih stranica, računa e-pošte ili korisnika kojima treba blokirati postavljanje ili dijeljenje na njihovoj platformi.
- Ovo je proces koji mogu podržati alati i dodaci koji ovaj posao mogu izvršiti automatski
- Spomenuti alati podržavaju moderatore u stručnom izvršavanju ovog zadatka, kao i redovito ažuriranje njihove crne liste / blok-liste dodavanjem ili uklanjanjem pojmova, računa, izvora sadržaja

A red umbrella is shown in the center, tilted slightly to the right. Above the umbrella, a shower of red icons is falling. These icons include symbols for spam (a crossed-out envelope), adult content (a heart with a slash), and other prohibited items (a triangle with an exclamation mark, a crossed-out speech bubble, and various symbols like 'XXX', '+18', and 'SPAM'). The background of the slide is dark grey.

Drugim riječima, moderatori sadržaja mogu sastaviti arhivu riječi ili izraza koji nisu dopušteni za njihovu platformu, kao i popis internetskih stranica, računa e-pošte ili korisnika kojima treba blokirati postavljanje ili dijeljenje na njihovoj platformi. Ovaj proces mogu podržati alati i dodaci koji ovaj posao mogu izvršiti automatski, podržavajući moderatore u stručnom izvršavanju ovog zadatka i održavajući njihove standarde na njihovim platformama ili računima, kao i redovito ažuriranje njihove crne liste / liste blokiranja dodavanjem ili uklanjanjem pojmova, računa, izvora sadržaja itd.



Danas se oslanjamo na **Internet mediji** da biste vijesti dobivali puno češće od tradicionalnih izvora vijesti.

**Internet forumi** dopustite nam da podijelimo znanje, komentiramo, poručujemo ili raspravljamo o određenoj temi s drugim ljudima sa istim zanimanjem.

**Društvene mreže ili web stranice**, kao što su Facebook, Instagram, Twitter, YouTube, Tumblr, LinkedIn, Snapchat, Quora, Reddit, Pinterest, koje naširoko koriste ljudi širom svijeta i omogućuju nam da tamo pronađemo i podijelimo bilo kakve vijesti.

**Podcast** omogućuje bilo kome da podijeli svoje znanje i komunicira sa svijetom kroz seriju zvukova usredotočenih na određenu temu ili temu.

## **Alati za moderiranje sadržaja**

Povećana potreba za moderiranjem sadržaja podijeljenog putem mrežnih platformi dovela je do razvoja širokog spektra alata i platformi za podršku moderatorima i administratorima u tom procesu.

Ti alati omogućuju sve od filtriranja vođenih umjetnom inteligencijom do jednostavnih dodataka koji im pomažu u održavanju određene kontrole i nadzora.

I na kraju, platforma može odlučiti da nema niti jedan oblik moderiranja ili nadgledanja. To se ne savjetuje jer nedostatak moderiranja u praksi znači da vlasnici (upravitelji) platforme nemaju kontrolu nad sadržajem koji bi mogao dovesti do toga da postane domaćin neprimjerenog, prijetećeg ili ilegalnog materijala i pogleda .

## **Alati za moderiranje sadržaja**

### **Alati za moderiranje komentara za blogere**

#### Disqus

Dodatak koji administratorima omogućuje pregled i moderiranje komentara, kao i stvaranje i upravljanje popisima zabrana korisnika, filtrima riječi, kontrolama neželjene pošte.

#### Komentari na Facebooku

Dodatak koji svojim korisnicima omogućuje povezivanje njihovih komentara sa njihovim Facebook profilima, a moderatorima omogućuje moderiranje komentara putem aplikacije Facebook ili putem preglednika i primjene skupnih radnji na komentare.

#### Disqus

Discus je alat koji se lako može instalirati na blog putem padajućeg koda ili kao dodatak i omogućava administratorima da pregledaju i moderiraju komentare na članke putem jedne nadzorne ploče. Nadzorna ploča također im omogućuje stvaranje, između ostalih, popisa zabrana korisnika, filtara riječi, kontrola neželjene pošte.

#### Komentari na Facebooku

Blogovi mogu koristiti dodatak za komentare na Facebooku koji omogućava svojim korisnicima da povežu svoje komentare sa svojim Facebook profilima, bez potrebe da se registriraju i prijave na navedeni blog. Ovaj alat omogućuje moderatorima da organiziraju komentare prema objavljenom vremenu ili u skladu s njihovim angažmanom. Što je najvažnije, daju im mogućnost moderiranja komentara putem aplikacije Facebook ili putem preglednika i primjene skupnih radnji na komentare. Oni također mogu donijeti odluku hoće li ostaviti kao javni ili sakriti komentare koji su označeni od strane Facebooka ili drugih korisnika (na primjer zbog uvredljivosti ili nepristojnosti). Napokon, ako moderiranje web stranice / bloga obavlja tim, moderatori također mogu dodijeliti određene komentare različitim članovima tima.



## **Alati za moderiranje sadržaja**

### **Alati za moderiranje komentara za blogere**

#### IntenseDebate

Omogućuje moderatorima da pregledaju komentare i odgovore na njih putem e-pošte, kao i da podijele zadatak moderiranja unutar tima administratora. Oni također mogu primijeniti filtre, pretraživati i / ili automatski brisati komentare prema ključnoj riječi, IP adresi ili adresi e-pošte i po potrebi zabraniti korisnicima.

#### Livefyre

Livefyre omogućuje moderatorima da pregledaju komentare prije nego što se objave, dopušta korisnicima da ih uređuju dok stvaraju popis zabrane korisnika i pravila za automatsko upravljanje komentarima

#### IntenseDebate

IntenseDebate je alat koji se može sinkronizirati s raznim platformama web stranica kao što su Wordpress, Tumblr ili Blogger. Omogućuje moderatorima da pregledaju komentare i odgovore na njih putem e-pošte, kao i da podijele zadatak moderiranja unutar tima administratora. Što je najvažnije, njegove značajke pružaju mogućnost moderatorima da filtriraju, pretražuju i / ili automatski brišu komentare prema ključnoj riječi, IP adresi ili adresi e-pošte i po potrebi zabranjuju korisnicima.

#### Livefyre

Livefyre sadrži razne prilagodljive postavke koje pomažu automatskom provođenju postupka moderiranja. Moderatorima imaju mogućnost pregleda komentara prije nego što se objave, omogućuju korisnicima da ih uređuju, kao i mogućnost stvaranja popisa zabrana koji određenim korisnicima onemogućuju komentiranje. Osim toga, oni mogu postaviti pravila za komentare, tako da ako se primjerice određeni broj korisnika označi nekim dijelom sadržaja, on se automatski briše. Isto se odnosi i na mrežni filter vulgarnosti.

## **Alati za moderiranje sadržaja**

### **Alati za rješavanje problema zlostavljanja na društvenim mrežama**

#### Privremeno isključivanje

Pri primanju zlouporabe na mreži putem Twittera, poželjno je da moderatori privremeno iskljue umjesto da blokiraju dotične račune. Ovaj pristup ublažava učinak zlouporabe jer ciljani račun ne prima obavijesti s privremeno isključenog računa.

#### Blokiranje

Ovo se smatra posljednjom radnjom protiv računa koji uporno šalju neželjenu poštu ili uvredljiv sadržaj.

## **Alati za moderiranje sadržaja**

**Alati za rješavanje problema zlostavljanja na društvenim mrežama**

### Prijavljivanje

Moderatori Twitteru obično prijavljuju tweetove ili račune koji šire potencijalno vjerodostojne i neposredne prijetnje ili sadrže nasilne slike.

## **Alati za moderiranje sadržaja**

### **Alati za rješavanje problema zlostavljanja na društvenim mrežama**

#### Izbriši komentar

Facebook pruža mogućnost administratorima stranica da izbrišu komentare koje mogu smatrati uvredljivima, prijetnjama ili uvredama.

#### Sakrij komentar

Moderatori mogu odlučiti samo sakriti uvredljivi komentar ispod objave. Međutim, ovo se smatra manje učinkovitim od brisanja jer to znači da dotični komentar ostaje vidljiv korisniku i njegovim prijateljima.

## **Alati za moderiranje sadržaja**

### **Alati za rješavanje problema zlostavljanja na društvenim mrežama**

#### Zabraniti korisniku pristup stranici

Ako korisnik opetovano objavljuje komentare koji krše standarde stranice, narušavajući vrijednosti zdrave rasprave, tada se savjetuje da se njegov / njezin profil zabranjuje.

#### Onemogućite / isključite komentare

Ova značajka dostupna je samo na video objavama i moderatori je odabiru kad smatraju da nemaju kapacitet za uspješno praćenje protoka komentara na videozapisu ili streamu uživo.

## **Alati za moderiranje sadržaja**

### **Alati za rješavanje problema zlostavljanja na društvenim mrežama**

#### Blokiranje riječi

Moderatori mogu pojačati filter psovki na svojoj stranici zabranom određenih riječi, čime osiguravaju da se komentari koji ih uključuju ne objavljuju na njihovoj stranici.

#### Prijavljivanje

Ako smatra da komentar / objava krši standarde Facebooka, moderator može prijaviti korisnika ili stranicu koja ga je napisala.

## **Alati za moderiranje sadržaja**

### **Alati / platforme za automatizirano moderiranje sadržaja**

#### Akismet

Akismet je dodatak koji se prvenstveno fokusira na blokiranje neželjene pošte. Također se može koristiti kao alat za skeniranje za nadgledanje postova i stranica, kao i za provjeru komentara kako bi se osiguralo da se zlonamjerni sadržaj ne pojavi na internetskoj stranici. Za više informacija:

<https://akismet.com/>

#### Utopia AI Moderator

Utopia AI Moderator potpuno je automatizirani alat za moderiranje posebno dizajniran za praćenje i iskorjenjivanje uvredljivih, lažnih i neželjenih sadržaja. Omogućuje administratorima da definiraju svoje politike moderiranja i može obrađivati sadržaj na više jezika. Za više informacija:

<https://utopiaanalytics.com/utopia-ai-moderator/>

#### Akismet

Akismet je dodatak koji se najbolje uklapa u platforme temeljene na WordPressu i prvenstveno se fokusira na blokiranje neželjene pošte. Također se može koristiti kao alat za skeniranje za nadgledanje postova i stranica, kao i za provjeru komentara kako bi se osiguralo da se zlonamjerni sadržaj ne pojavi na internetskoj stranici. Važno je napomenuti da je ovaj dodatak i jednostavan za upotrebu i besplatan.

Za više informacija: <https://akismet.com/>

#### Utopia AI Moderator

Utopia AI Moderator potpuno je automatizirani alat za moderiranje posebno dizajniran za praćenje i iskorjenjivanje uvredljivih, lažnih i neželjenih sadržaja. Alat moderira 100% dolaznog sadržaja i on ostaje u toku s učenjem dok djeluje ispitujući odluke o objavljivanju koje su ljudski moderatori donosili u prošlosti. Omogućuje administratorima da definiraju svoju politiku moderiranja, koju automatski slijede i mogu obrađivati sadržaj na više jezika, obrađivati veze i adrese. Utopia AI Moderator također je sposoban razumjeti kontekst sadržaja koji se procjenjuje, bez obzira na žargon i pravopis, dok nudi automatsko moderiranje videozapisa i slika.

Za više informacija: <https://utopiaanalytics.com/utopia-ai-moderator/>

## **Alati za moderiranje sadržaja**

### **Alati / platforme za automatizirano moderiranje sadržaja**

#### Implio

Alat koji podržava automatizirani i ručni sadržaj, kao i moderiranje komentara. Nudi ručno moderiranje putem sučelja za kontrolu, a moderatorima također omogućuje postavljanje širokog niza pravila automatizacije radi racionalizacije cijelog procesa.

Za više informacija: <https://besedo.com/implio-features/>

#### PicPurify

PicPurify je sučelje za programiranje aplikacija za moderiranje fotografija u stvarnom vremenu i dizajniran je za automatsko otkrivanje i filtriranje slika sa neželjenim sadržajem. U mogućnosti je prepoznati štetne elemente na slikama, a moderatori mogu postaviti parametre koje žele. Za više informacija:

<https://www.picpurify.com/>

#### Implio

Implio je alat koji se može iskoristiti i podržava automatizirani i ručni sadržaj, kao i moderiranje komentara. Nudi sučelje za ručno moderiranje za kontrolu postupka i omogućava korisnicima da postave svoje prilagođene filtre za lako ciljanje i uklanjanje neželjenog sadržaja s predvidljivim uzorcima. Uz to, omogućuje im postavljanje širokog niza pravila automatizacije radi racionalizacije cijelog procesa.

Za više informacija: <https://besedo.com/implio-features/>

#### PicPurify

PicPurify je sučelje za programiranje aplikacija za moderiranje fotografija u stvarnom vremenu i dizajniran je za automatsko otkrivanje i filtriranje slika sa neželjenim sadržajem. U mogućnosti je prepoznati štetne elemente na slikama kao što su golotinja, droga, mržnja i osigurati da se oni ne pojavljuju na platformi. Moderatori su u stanju stvoriti prilagođeni pristup koji odgovara njihovim potrebama i perspektivi, a alati djeluju 24 sata dnevno kako bi osigurali da njihovi zahtjevi budu ispunjeni.

Za više informacija: <https://www.picpurify.com/>



## **Alati za moderiranje sadržaja**

### **Alati / platforme za automatizirano moderiranje sadržaja**

#### WebPurify automatizirano inteligentno moderiranje

Alat koji nudi konstantnu zaštitu od rizika povezanih sa slikama koje generiraju korisnici otkrivanjem i uklanjanjem golotinje i drugog neprikladnog sadržaja u stvarnom vremenu. Za više informacija: <https://www.webpurify.com/photo-moderation/automated/>

#### Moderator sadržaja Azure

Moderator sadržaja Azure kognitivna je usluga koja nadgleda sadržaj u više oblika, poput tekstualnog, slikovnog i video sadržaja, provjeravajući i primjenjujući odgovarajuće oznake (zastavice) za materijale koji mogu biti uvredljivi, rizični ili na drugi način nepoželjni.

Za više informacija: <https://docs.microsoft.com/en-in/azure/cognitive-services/content-moderator/overview>

#### WebPurify automatizirano inteligentno moderiranje

Usluga Automatiziranog inteligentnog moderiranja (AIM) WebPurify nudi neprekidnu zaštitu od rizika povezanih sa slikama koje generiraju korisnici otkrivanjem i uklanjanjem golotinje i drugog neprikladnog sadržaja u stvarnom vremenu. Alati mogu otkriti slike koje sadrže golotinu, alkohol, droge, uvredljive geste i simbole i tekstove mržnje kako bi osigurali da sadržaj na platformi bude "siguran" od takvog uvredljivog ili lažnog sadržaja. Štoviše, nudi prilagođene moderacijske modele koji se mogu uklopiti i najbolje služiti potrebama svake platforme.

Za više informacija: <https://www.webpurify.com/photo-moderation/automated/>

#### Moderator sadržaja Azure

Azure Content Moderator je kognitivna usluga koja nadgleda sadržaj u više oblika, poput tekstualnog, slikovnog i video sadržaja, provjeravajući i primjenjujući odgovarajuće oznake (zastavice) za materijale koji mogu biti uvredljivi, rizični ili na drugi način nepoželjni. Usluga moderatora sadržaja također uključuje alat za pregled zasnovan na webu, koji nudi recenzije sadržaja koje obrađuju ljudski moderatori. Kombinirajući rad usluge s timovima za ljudski pregled, moderatori platforme mogu postići pravu ravnotežu između učinkovitosti i točnosti. Alat za pregled također nudi korisnički prilagodljiv prednji kraj za nekoliko resursa Moderatora sadržaja.

Za više informacija: <https://docs.microsoft.com/en-in/azure/cognitive-services/content-moderator/overview>



Intelektualno vlasništvo

Izvor videozapisa: <https://www.youtube.com/watch?v=UqZJPuyK9VY>

### **Potreba za moderiranjem sadržaja na društvenim mrežama**

- Platforme društvenih medija postale su najvažniji izvori sadržaja i informacija.
- Iako je ovo donijelo revoluciju u načinu na koji se vijesti i informacije konzumiraju, dijele i o kojima se raspravlja, platforme društvenih medija također su osigurale vakuum za širenje dezinformacija, mržnje, internetskog maltretiranja i zlonamjernog sadržaja.
- Slijedom toga, sve je veća potražnja za time da platforme društvenih medija poduzmu mjere protiv takvih oblika sadržaja, što ih navodi na poduzimanje koraka za praćenje i moderiranje.

Platforme društvenih medija postale su najvažniji izvori sadržaja i informacija. Veliki dio internetskih korisnika oslanja se na društvene medije kako bi pristupili informacijama, vijestima i gledištima, kao i da bi stupili u interakciju sa svojim vršnjacima. Iako je ovo donijelo revoluciju u načinu na koji se vijesti i informacije konzumiraju, dijele i o kojima se raspravlja, platforme društvenih medija također su osigurale vakuum za širenje dezinformacija, mržnje, internetskog maltretiranja i zlonamjernog sadržaja.

Slijedom toga, sve je veća potražnja za time da platforme društvenih medija poduzmu mjere protiv takvih oblika sadržaja, što ih navodi na poduzimanje koraka za praćenje i moderiranje.

### **Odgovor platformi društvenih medija**

- Platforme društvenih medija razvile su standarde zajednice koji ocrtavaju što je prihvatljivo, a što nije
- Oni koriste razrađene strategije moderiranja sadržaja koje uključuju ljudske moderatore, kao i automatizirane sustave kako bi osigurali da su njihovi korisnici zaštićeni u mjeri u kojoj je to moguće od nepristojnih sadržaja i zlostavljanja.
- Oni traže od korisnika da prijave sadržaj koji bi mogao uvrijediti ili krši pravila platforme

Platforme društvenih medija postale su najvažniji izvori sadržaja i informacija. Veliki dio internetskih korisnika oslanja se na društvene medije kako bi pristupili informacijama, vijestima i gledištima, kao i da bi stupili u interakciju sa svojim vršnjacima. Iako je ovo donijelo revoluciju u načinu na koji se vijesti i informacije konzumiraju, dijele i o kojima se raspravlja, platforme društvenih medija također su osigurale vakuum za širenje dezinformacija, mržnje, internetskog maltretiranja i zlonamjernog sadržaja.

Slijedom toga, sve je veća potražnja za time da platforme društvenih medija poduzmu mjere protiv takvih oblika sadržaja, što ih navodi na poduzimanje koraka za praćenje i moderiranje.


The background of the slide features a dark grey field filled with a dense network of white and blue icons. These icons represent various digital concepts: social media (Twitter bird, Facebook 'f', YouTube play button), communication (speech bubbles, mail, Wi-Fi signal), and technology (smartphone, laptop, cloud). The icons are interconnected by thin white lines, creating a web-like pattern. In the lower center, there is a faint, light grey silhouette of a smartphone. At the bottom of the slide, there is a solid orange horizontal bar.

Aktivnost br. 3: Grupna prezentacija o  
primjerima moderiranja na  
društvenim mrežama






Kada je u pitanju moderiranje sadržaja na društvenim mrežama, dolazi u pitanje sloboda govora i strah od cenzure ili kontrole promicanja podijeljenih mišljenja. Pogledajte ovaj videozapis kao uvod u raspravu koju možete voditi unutar grupe



**Hvala vam!**



If you exchange information internationally, you must strengthen data protection. Those are two sides of the same coin.

— *Gijs de Vries* —

Hvala!





Co-funded by the  
Erasmus+ Programme  
of the European Union

Potporna Europske komisije za izradu ove publikacije ne predstavlja  
odobrenje sadržaja, koji odražava samo stavove autora te se Komisiju  
ne može smatrati odgovornom za bilo kakvu uporabu koja može  
proizaći iz informacija u tekstu.

2019-1-DE02-KA204-006115

