

# Výročný report monitoringu nenávistných prejavov 2025

Február 2026

digiQ

## Zhrnutie

Online svet je integrálnou súčasťou spoločenského diania. Popri priestore pre tvorivosť, sebakvadenie a spájanie ľudí zároveň umožňuje šírenie a často aj zosilňovanie nenávistných naratívov prítomných v offline svete. Nenávistné prejavy online majú rovnako reálne dôsledky pre dôstojnosť, mentálne zdravie a bezpečie používateľov a nemožno ich redukovať na výlučne virtuálny problém.

Monitoring digiQ sa počas roka 2025 zameriaval na verejne dostupný obsah na štyroch najpoužívanějších sociálnych sieťach na Slovensku – Facebooku, Instagrame, YouTube a TikToku – s cieľom zhodnotiť v praxi fungovanie mechanizmov ochrany používateľov pred nenávistným obsahom. Dohromady sme analyzovali a nahlásili viac ako 700 nenávistných prejavov. Viac ako 80 % z nich malo charakter potenciálne nezákonného obsahu. Dominantnými boli výzvy na násilie a nenávisť, vrátane dehumanizačných naratívov aj podnecovania genocídy. Nenávistné prejavy sa najčastejšie zameriavali na LGBTI+ osoby, Rómov a utečencov, či ľudí s migračným pozadím, čo odráža pretrvávajúcu mieru stigmatizácie a zvýšené riziká, ktorým tieto skupiny čelia v online aj offline prostredí.



Report odhaľuje výrazné rozdiely v tom, ako jednotlivé platformy reagujú na nahlásenia zo strany používateľov. Výsledok nahlásenia sa ukazuje byť do veľkej miery ovplyvnený použitým nahlasovacím kanálom, nie závažnosťou obsahu. Zatiaľ čo špecializované nahlasovacie kanály pre expertné organizácie dosahujú vysokú mieru účinnosti, skúsenosť bežných používateľov je často sprevádzaná dlhými oneskoreniami alebo absenciou akejkoľvek odpovede, a to aj pri nahláseniach obsahujúcich explicitné výzvy na vážne formy násilia.

V kontexte európskeho Aktu o digitálnych službách prispievajú výsledky monitoringu k hodnoteniu praktického fungovania ochranných mechanizmov na sociálnych sieťach, identifikujú ich silné a slabé stránky a vytvárajú podklad pre ďalšie analytické a advokačné kroky.



## Online svet nie je oddelený od reality

Digitálne služby sa stali infraštruktúrou verejnej diskusie a online platformy dnes supľujú rolu verejných fór, médií aj komunitných priestorov. Pre veľkú časť populácie predstavujú primárny zdroj informácií, sociálnych kontaktov, diskusie o verejnom dianí aj zábavy. Obsah, ktorý sa v tomto prostredí šíri, má preto priamy vplyv na spoločenské normy aj na pocit bezpečia v spoločnosti. Dehumanizujúce a útočné naratívy, vrátane tých, ktoré propagujú násilie, nepôsobia len na konkrétne obete, ale aj na široké publikum, ktoré im je často vystavené nechcene.

Skúsenosti z posledných rokov – vrátane útokov motivovaných nenávisťou na Slovensku – ukazujú, že **hranica medzi online a offline násilím je čoraz priepustnejšia**. V roku 2022 sa odohral teroristický útok pred queer podnikom Tepláreň v Bratislave, kde útočník zastrelil dve osoby a jednu vážne zranil. Strelec sa zradikalizoval v online komunitách spájaných s militantným akcelerationizmom, kde tiež získal návody na prípravu útoku (Diko & Kőváry Sólmos, 2025). Začiatkom roka 2025 sa zase odohral útok na gymnázium v Spišskej Starej Vsi, pri ktorom zahynula zástupkyňa školy a študentka. Za útočnickovou motiváciou mal byť aj jeho nenávistný vzťah k ženám (Hlavanda Gogová, 2025); mizogýnné naratívy pritom v online priestore dlhodobo silnejú.

Digitálne platformy vytvárajú priestor, kde získava obsah stimulujúci silné emócie vyššiu viditeľnosť a vďaka algoritmom sa rýchlo šíri ďalej. Aj v dôsledku toho zohrávajú platformy **významnú rolu v procesoch radikalizácie a normalizácie extrémnych naratívov** (Castaño-Pulgarín et al., 2021; Lupu et al., 2023). Monitoring nenávistných prejavov tak má význam aj z pohľadu prevencie širších spoločenských a bezpečnostných rizík.

Monitoring digiQ plní dve funkcie. Na jednej strane má ochranný rozmer – identifikovaný obsah je nahlasovaný platformám s cieľom, aby ho v prípade, že porušuje pravidlá platformy alebo zákon, odstránili, či inak obmedzili. Na druhej strane ide o **kontrolný mechanizmus, ktorý umožňuje hodnotiť, ako si platformy v praxi plnia svoje záväzky** voči používateľom.



## Čo je nenávistný prejav?

V digiQ pracujeme s medzinárodne rešpektovanou definíciou Rady Európy (Všeobecné odporúčanie č. 15 Európskej komisie proti rasizmu a intolerancii), ktorá ako nenávistné prejavy vníma obhajovanie, podporu alebo podnecovanie znevažovania, nenávisti alebo hanobenia osoby alebo skupiny osôb, ako aj obťažovanie, urážanie, vytváranie negatívnych stereotypov, stigmatizáciu, vyhrážanie sa a odôvodňovanie všetkých týchto foriem prejavu identitárnymi znakmi, akými sú rasa, farba pleti, národnostný alebo etnický pôvod, zdravotné postihnutie, jazyk, náboženstvo, pohlavie, rod, pohlavná identita, sexuálna orientácia a iné osobné charakteristiky (Rada Európy, 2025). **Nenávistné prejavy môžu, ale nemusia porušovať zákon.**

V digiQ sa zameriavame na nezákonné prejavy, teda prejavy, ktoré môžu mimo iného napĺňať skutkové podstaty trestných činov podľa slovenského trestného zákona. Popri nezákonných prejavoch sledujeme v obmedzenej miere aj ubližujúce prejavy, ktoré neporušujú zákon, no sú v rozpore s komunitnými pravidlami platforiem, ktoré sa takýmito pravidlami snažia chrániť svojich používateľov. Tento typ obsahu využívame najmä na testovanie funkčnosti samoregulačných mechanizmov platforiem.

## Ako pracujeme s dátami

Sociálne siete monitorujeme organicky a prevažne manuálne, čo nám umožňuje zachytiť kontext, jazykové nuansy a dynamiku diskusií, ktoré by inak zostali skryté. Tento prístup má svoje limity. Dáta vzhľadom na obmedzenejšiu vzorku nepredstavujú štatisticky presný obraz online nenávisti na Slovensku.

Počas roka 2025 sme analyzovali **706 prípadov nenávistných prejavov** na štyroch najpopulárnejších sociálnych sieťach na Slovensku, na ktorých bolo možné sledovať verejne dostupný obsah na verejných stránkach, kanáloch alebo komunitách. **Monitorované neboli uzavreté profily ani neverejné skupiny, komunity či kanály, rovnako ako neverejná komunikácia.**

Nenávistné prejavy boli identifikované organicky pri monitorovaní obsahu, najmä na veľkých a populárnych kanáloch s výrazným dosahom na slovenské publikum. Každý prejav prešiel vyhodnotením na základe kvalitatívnych kritérií na identifikáciu nenávistných prejavov porušujúcich pravidlá komunity, ako aj na identifikáciu potenciálne nezákonných prejavov podľa platnej slovenskej legislatívy.

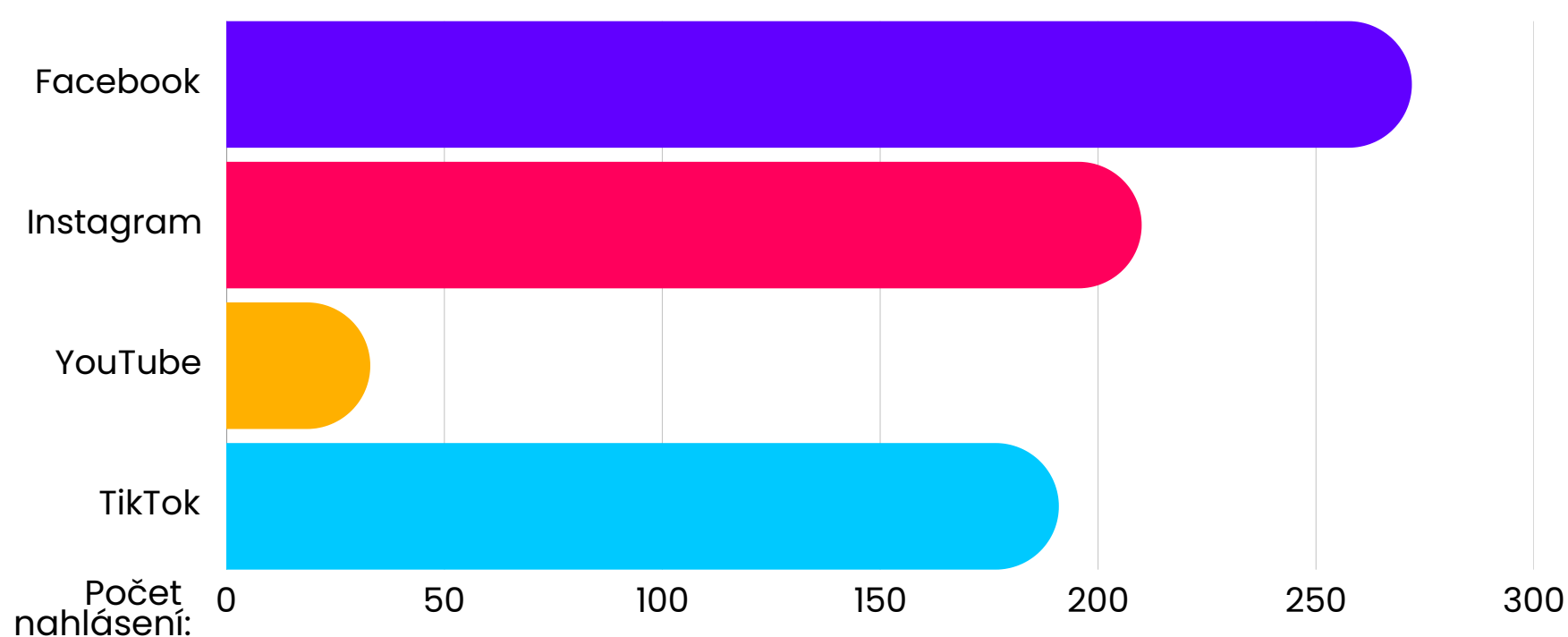
Všetky prípady boli individuálne nahlásené sociálnej sieti, následne boli hodnotené reakcie platforiem. Sledované boli najmä tieto aspekty:

- **responzivnosť:** či platformy nahlásenie spracovali a oboznámili nahlasovateľa s podniknutými krokmi,
- **rýchlosť reakcie:** ako rýchlo platformy nahlásenia vyhodnotili a oboznámili o tom nahlasovateľa,
- **náprava:** či platformy obsah odstránili, prípadne inak zamedzili prístup k nemu.



## Povaha dát monitoringu

Distribúciu vyše 700 nahlásených nenávistných prejavov na jednotlivých platformách je vidieť na grafe nižšie. **83 % nahlásení** z celkového objemu bolo **vyhodnotených ako potenciálne nezákonný obsah**. Zvyšných 17 % prípadov tvorili ubližujúce, no zákonné prejavy namierené voči jednotlivcom alebo skupinám na základe ich identitárnych znakov, ktoré boli platformám nahlásené ako porušenia ich pravidiel komunity.



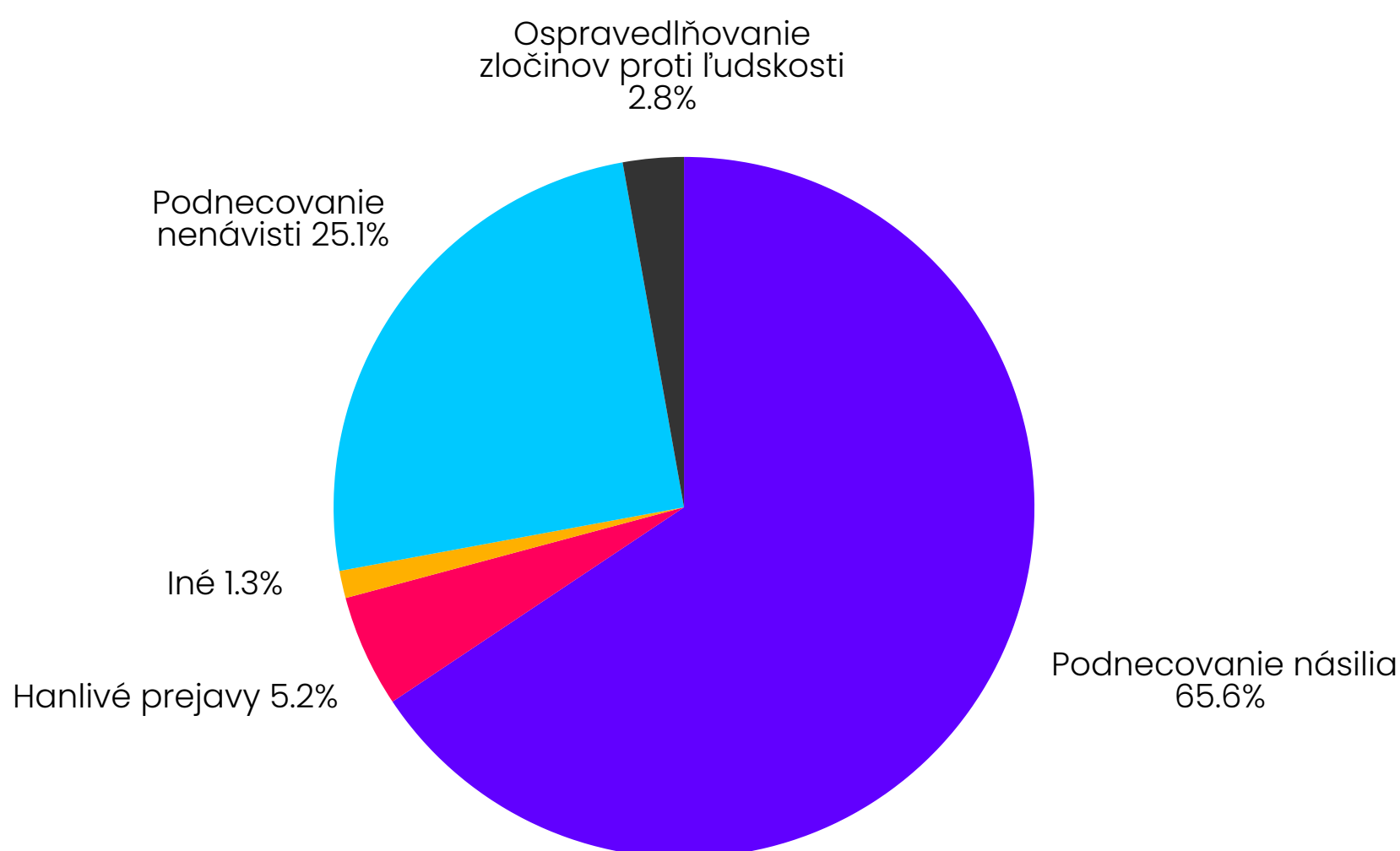
**Graf 1:** Distribúcia nenávistných prejavov na jednotlivých platformách v datasete digiQ

Rozdiely v objeme identifikovaných nenávistných prejavov neodzrkadľujú s presnosťou pomer výskytu nenávistného obsahu naprieč platformami, keďže monitoring bol realizovaný bez systematického využitia softvérových nástrojov. Tieto rozdiely však môžu do určitej miery poukazovať na odlišné vzorce správania používateľov v rámci jednotlivých digitálnych prostredí.

Významnú úlohu zohráva architektúra platforiem, teda spôsob, akým sú navrhnuté diskusné priestory a mechanizmy interakcie. V prípade TikToku možno hypotetizovať, že krátky, dynamický a prevažne vizuálny obsah znižuje priestor pre kontextualizáciu a podporuje rýchle, emotívne reakcie, ktoré môžu zahŕňať aj nenávistné prejavy. Naopak, pri YouTube sa zdá, že dizajn platformy podporuje skôr pasívnu konzumáciu obsahu než intenzívnu interakciu v diskusiách. YouTube má pritom na Slovensku podľa dostupných odhadov takmer štyri milióny používateľov – teda výrazne viac než TikTok, ktorý tu zatiaľ oslovuje vekové skupiny nad 25 rokov len v menšej miere (digiQ, 2025a; DataReportal, 2025). Zároveň, nižší počet nahlásených prejavov digiQ na tejto platforme odzrkadľuje aj dlhodobu nefunkčnosť nahlasovacích mechanizmov YouTube – platforma od roku 2023 neodstránila žiaden z nenávistných prejavov, ktoré sme nahlásili (digiQ, 2025b).

## Rozsah a formy nenávisti

Najväčší podiel datasetu, takmer **dve tretiny** prípadov, obsahoval výzvy na násilie, pričom zastúpené boli aj genocídne naratívy. Ďalších vyše **25 %** prípadov predstavovalo podnecovanie nenávisti bez výslovných zmienok o násilí. Menšia časť, **2,8 %** prípadov, zahŕňala naratívy ospravedlňujúce alebo glorifikujúce vojnové zločiny a zločiny proti ľudskosti. Rasistické, xenofóbne alebo sexistické urážky, ktoré boli nahlásené ako ubližujúci obsah porušujúci pravidlá komunity platforiem, tvorili dohromady asi **5 %** datasetu.

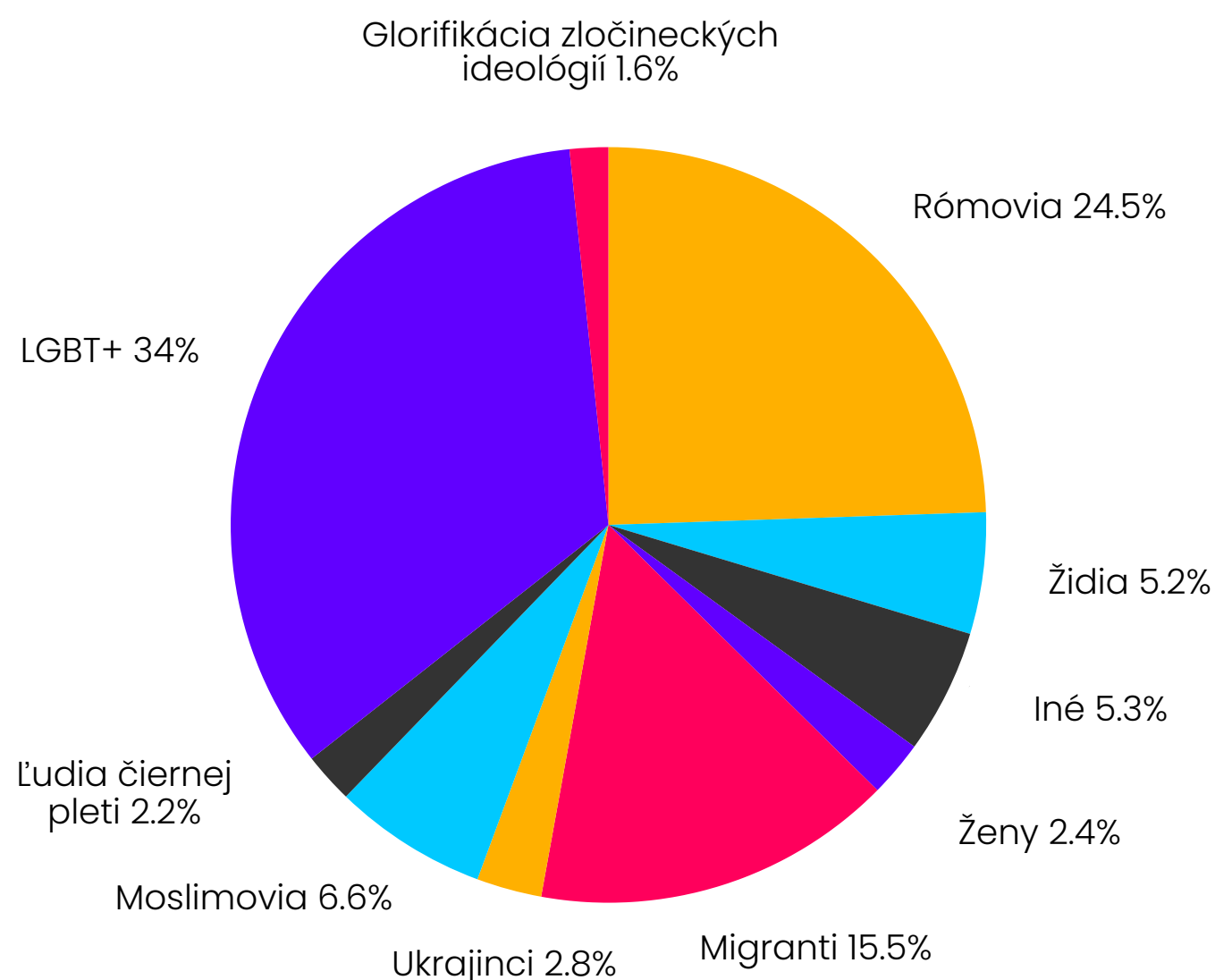


**Graf 2:** Formy nenávistných prejavov v datasete digiQ

## Cieľové skupiny a opakujúce sa vzorce

Nenávistné prejavy sa sústreďujú najmä na skupiny, ktoré sú už v offline prostredí vystavené sociálnemu vylúčeniu, dlhodobej stigmatizácii, a častokrát aj nešťastnej politickej instrumentalizácii. Online prostredie tento tlak vďaka vysokému dosahu, rýchlemu šíreniu, a napríklad aj pocitu anonymity používateľov, často zosilňuje.





**Graf 3:** Cieľové skupiny nenávisti na Slovensku

### Ako majú platformy reagovať?

Podľa Aktu o digitálnych službách (DSA) majú takzvané veľmi veľké online platformy (VLOPs), kam patria všetky monitorované platformy v našom monitoringu, osobitné povinnosti pri spracúvaní nahlásení nenávistných prejavov (Európska únia, 2022). Povinnosti platformám vznikajú nielen, ak ide o nezákonný obsah, ale regulujú tiež obsah porušujúci pravidlá komunity.

VLOPs sú povinné zaviesť **ľahko dostupné a používateľsky prívetivé mechanizmy nahlasovania** (článok 16 DSA), ktoré umožňujú používateľom a oprávneným subjektom oznámiť podozrenie na nezákonný obsah, mimo iného na nezákonné nenávistné prejavy. Po prijatí oznámenia musia platformy konať bez zbytočného odkladu, individuálne obsah posúdiť a prijať primerané opatrenia (napr. odstránenie alebo obmedzenie prístupu). Rozhodnutia musia byť odôvodnené a transparentné. Oznamovateľ musí byť informovaný o výsledku, dôvodoch a zvolených prostriedkoch nápravy (články 16 a 17 DSA).

DSA zároveň vyžaduje, aby platformy **dôsledne a transparentne uplatňovali svoje pravidlá komunity** (článok 14). VLOPs musia zabezpečiť, že spracovanie nahlásení porušení týchto pravidiel je založené na jasných pravidlách a rovnako informovať používateľa o výsledku, dôvodoch a prostriedkoch nápravy.

Ako subjekty so systémovým vplyvom majú VLOPs navyše povinnosť **posudzovať a zmierňovať systémové riziká**, vrátane šírenia nenávistných prejavov a ich dopadov na základné práva (články 34 a 35 DSA). To zahŕňa aj povinnosť zabezpečiť, aby mechanizmy nahlasovania a moderácie neboli len formálne dostupné, ale reálne účinné v praxi, čomu má zodpovedať aj rýchlosť a kvalita spracovania nahlásení. Monitorované platformy sa v Kódexe správania pre boj proti nezákonným prejavom online, ktorý dopĺňa legislatívu DSA, zaväzujú usilovať sa o to, aby bola väčšina nahláseného obsahu spracovaná do 24 hodín od nahlásenia (Európska únia, 2025).

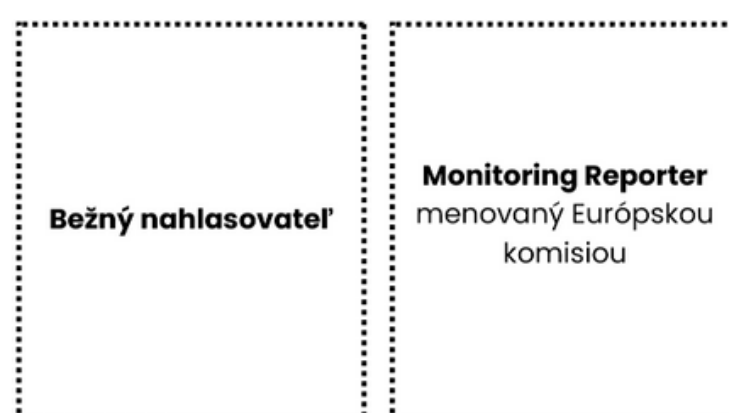
### ... aké prostriedky nahlasovania sú teda dostupné?

**Pravidlá komunity** predstavujú vnútorné pravidlá, ktoré si platformy sami vytvárajú, aby ochránili používateľov pred nebezpečným, ubližujúcim alebo toxickým obsahom. Pravidlá komunity nie sú prispôbolené lokálnemu kontextu. Zo svojej podstaty však bývajú častokrát prísnejšie ako národné zákony – zakazujú napríklad používanie veľkého množstva hanlivých označení, škodlivých stereotypov a iných foriem výsmechu.

Vďaka DSA môžu v súčasnosti používatelia samostatne nahlasovať aj nezákonný obsah, resp. nezákonné nenávistné prejavy, o ktorých sa domnievajú, že porušujú zákony v ich krajine. Nahlasovať tak môžu aj prejavy, ktoré sú lokálne špecifické, častokrát spojené s historickým kontextom krajiny. Článok 16 DSA predstavuje mechanizmus oznamovania a prijímania opatrení (**mechanizmus Notice and Action**), na základe ktorého majú platformy vytvárať kanály pre nahlasovanie nezákonného obsahu dostupné všetkým používateľom. Kanály umožňujú nahlasovateľovi popísať kontext a okolnosti prejavu a označiť, ktoré zákonné úpravy porušuje – čo môže napomôcť efektívnejšiemu spracovaniu na strane platformy.

Špecializované kanály pre nahlasovanie nezákonných nenávistných prejavov sú zase dostupné *dôveryhodným nahlasovateľom*, ktorých menujú národní koordinátori DSA a ktorých nahlásenia spracovávajú platformy prioritne. V kontexte DSA vznikol tiež štatút *Monitoring Reporter*, ktorý expertným organizáciám udeľuje Európska komisia. Monitoring Reporters participujú na každoročných monitorovacích cvičeniach Európskej komisie s cieľom testovať, ako VLOPs plnia svoje záväzky pri spracovávaní nahlásení.

Niektoré platformy v tomto kontexte iniciatívne vytvorili nahlasovacie **kanály, ktoré sú dostupné práve Monitoring Reporters**, a ktoré im, podobne ako v prípade dôveryhodných nahlasovateľov, umožňujú nahlasovať prioritne. Medzi platformami, ktoré sme v digiQ monitorovali v roku 2025, mali takýto kanál platformy spoločnosti Meta.

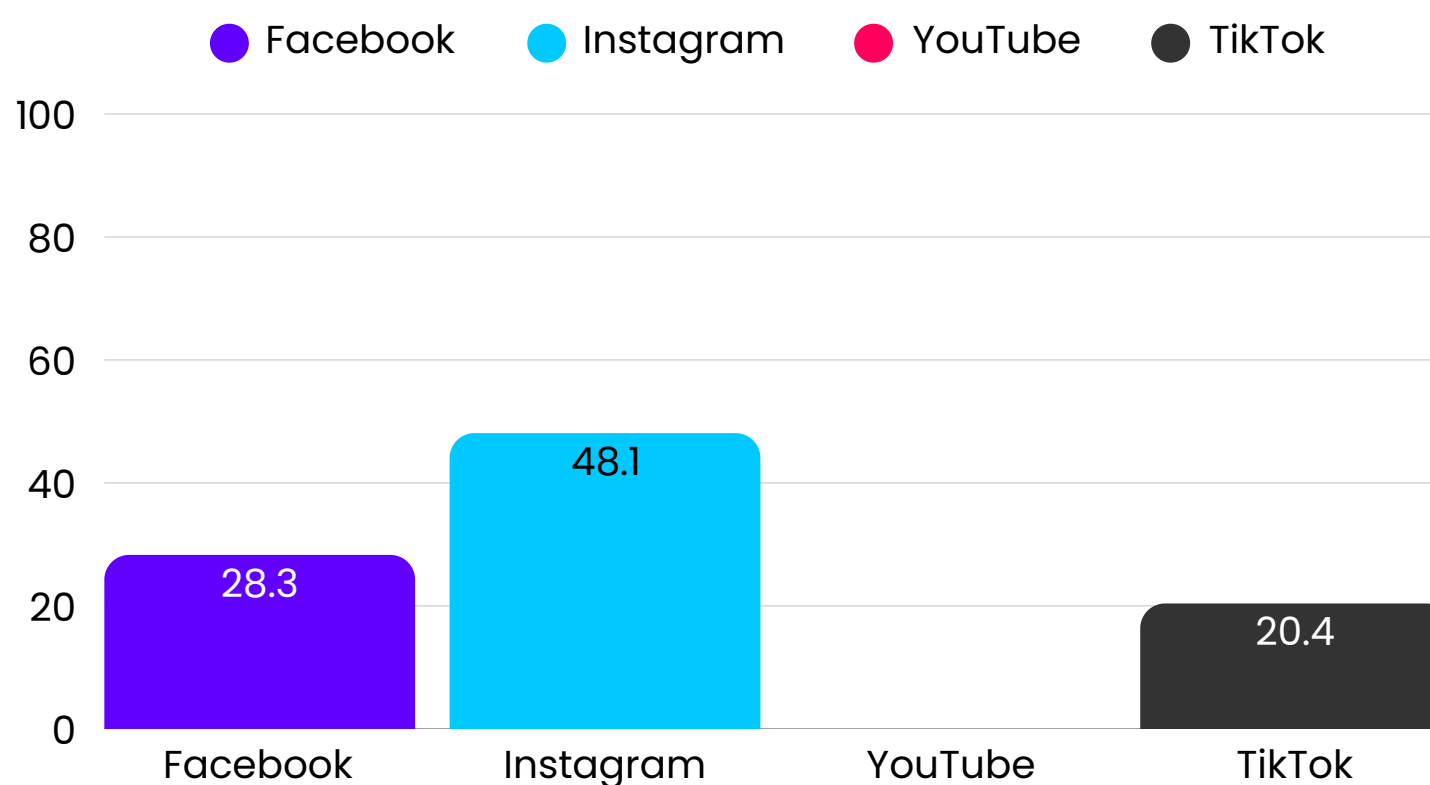


## Ako platformy reagujú v praxi

DSA stanovuje jasné pravidlá, ako by nahlasovanie malo fungovať. Ako však monitorované platformy postupujú v praxi? Zamerali sme sa na tri kľúčové premenné: celkovú **responzivnosť platformy, rýchlosť reakcie** na nahlásenia a **mieru odstraňovania** nahláseného obsahu. Iné formy nápravy, než je odstránenie, sme počas monitoringu zaznamenali len netradične. Zmieňované premenné poskytujú empirický obraz fungovania mechanizmov moderácie obsahu v kontexte povinností vyplývajúcich z DSA.

### Odstraňovanie obsahu: účinnosť závislá od kanála, nie od obsahu

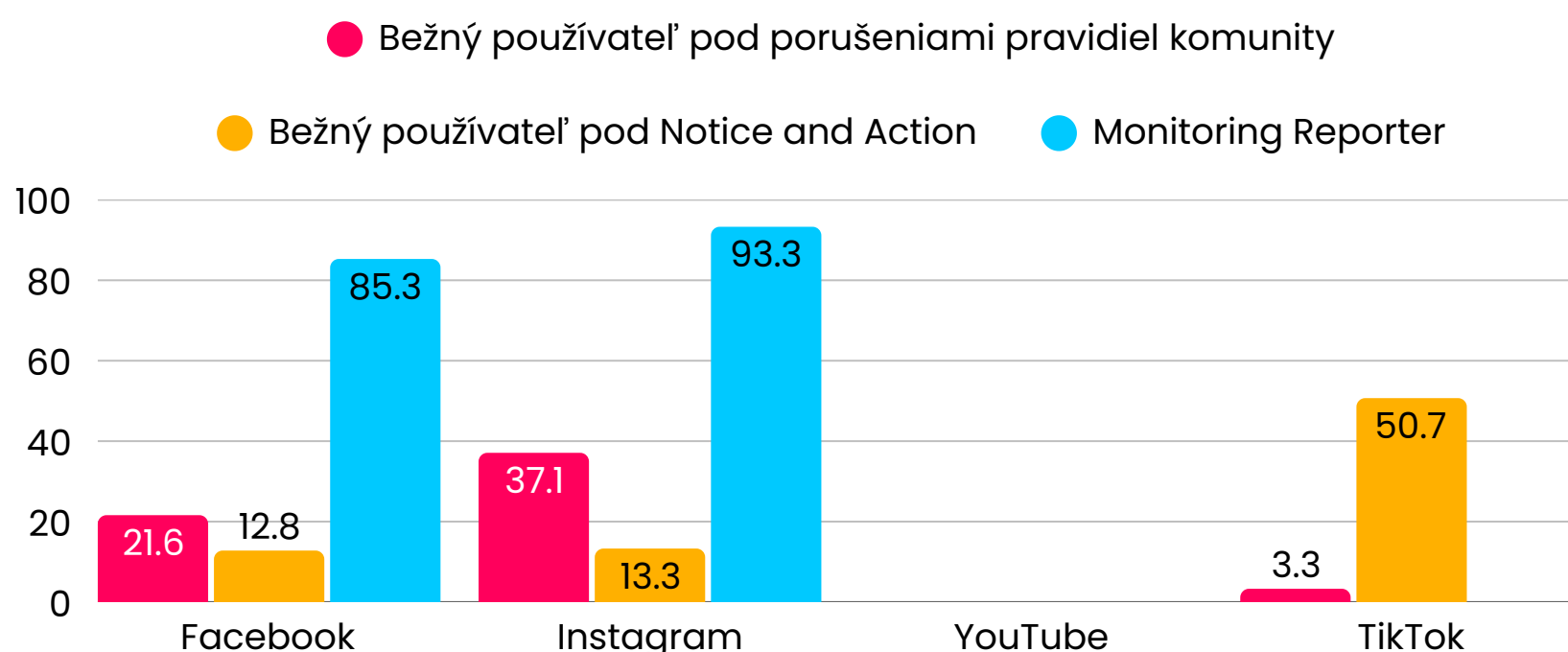
Miera odstraňovania sa naprieč platformami výrazne líšila. Najvyššiu mieru odstraňovania dosiahol Instagram, kde bolo odstránených **48,1 %** celkového nahláseného obsahu. Facebook odstránil **28,3 %** nahlásených prípadov, teda výrazne menej než Instagram, ktorý rovnako patrí pod spoločnosť Meta. TikTok dosiahol ešte nižšiu mieru odstránenia (**20,4 %**), zatiaľ čo YouTube počas roku neodstránil žiadny z nahlásených prípadov.



**Graf 4a:** Celková miera odstránenia nahláseného obsahu (vyjadrené v %)

**Rozdiely vo výsledkoch úzko súvisia s tým, akými nahlasovacími kanálmi platformy disponujú a ako k nim pristupujú.** Všetky monitorované platformy umožňujú ktorémukoľvek používateľovi nahlasovať porušenia pravidiel komunity, ako aj nezákonný obsah pod mechanizmom Notice and Action. Platformy spoločnosti Meta taktiež umožňujú nahlasovať nezákonný obsah expertným organizáciám so štatútom Monitoring Reporter prostredníctvom špecializovaného kanála. Nasledujúci graf zobrazuje mieru odstraňovania prostredníctvom týchto kanálov na jednotlivých platformách.





**Graf 4b:** Miera odstránenia pri nahlasovaní prostredníctvom jednotlivých kanálov. Zatiaľ čo platformy spoločnosti Meta umožňujú organizáciám so štatútom Monitoring Reporter nahlasovať cez špecializovaný kanál, YouTube ani TikTok takéto kanály neposkytujú.

Na **Facebooku** je rozdiel medzi kanálmi obzvlášť výrazný. V prípade špecializovaného kanála pre Monitoring Reporters dosahovala miera odstránenia **85,3 %**, zatiaľ čo pri kanáli pod mechanizmom Notice and Action pre bežných používateľov len **12,8 %**. V oboch prípadoch bol pritom nahlasovaný obdobný nezákonný obsah. Pri nahlasovaní porušenia pravidiel komunity z pozície bežného používateľa bola miera odstránenia **21,6 %**, no významná časť zásahov prišla až po tom, čo sme obsah nahlásili druhýkrát – po prvotne zamietavej spätnej väzbe zo strany platformy.

**Instagram** taktiež dosahoval vysokú mieru odstraňovania v špecializovanom kanáli pre Monitoring Reporters (**93,3 %**), no vykazoval relatívne lepšie výsledky aj pri nahlasovaní porušenia pravidiel komunity (**37,1 %**). Avšak pri kanáli pod mechanizmom Notice and Action zostala miera odstránenia nezákonného obsahu veľmi nízka, s výsledkom **13,3 %**.

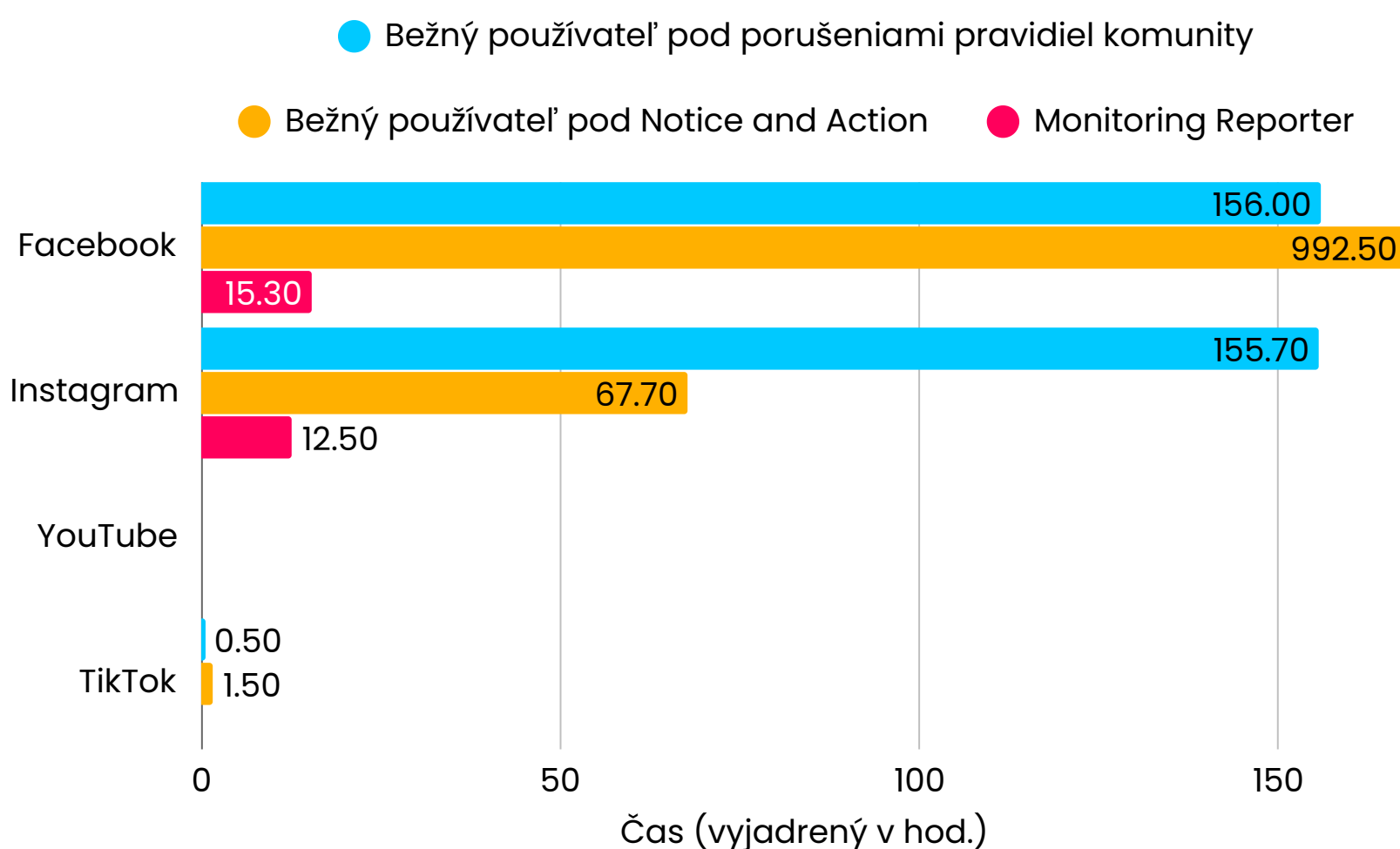
**TikTok** nemá k dispozícii špecializovaný kanál pre Monitoring Reporters, no rozdiely medzi jeho štandardnými mechanizmami sú výrazné. Pri nahlasovaní nezákonného obsahu pod mechanizmom Notice and Action dosahovala miera odstránenia **50,7 %**. Pri nahlasovaní porušenia pravidiel komunity dosahovala len **3,3 %** a obsah bol často odstránený až po opätovnom nahlásení. Napriek rýchlemu spracovaniu tak pri tomto kanáli dochádza k minimálnemu zásahu voči nepovolenému obsahu.

Pri porovnaní platforiem a kanálov sa ukazuje, že **závažnosť obsahu nemá systematický vplyv na mieru odstraňovania**. V rámci nahlasovania porušenia pravidiel komunity bola miera odstránenia ubližujúcich, ale zákonných prejavov približne **15 %**, zatiaľ čo pri obsahu vyhodnotenom ako potenciálne nezákonný dosahovala len mierne vyšších **20,1 %**. Ani pri nahláseniach obsahujúcich explicitné výzvy na násilie, vrátane ozbrojeného násillia, neprekročila miera odstránenia **21,3 %**. V praxi tak často dochádzalo k situáciám, kedy platformy odstránili menej závažné prejavy, no ignorovali nahlásenia obsahujúce výzvy na zabitie či mučenie.

## Rýchlosť reakcie ako kritický faktor ochrany: spracovanie a spätná väzba

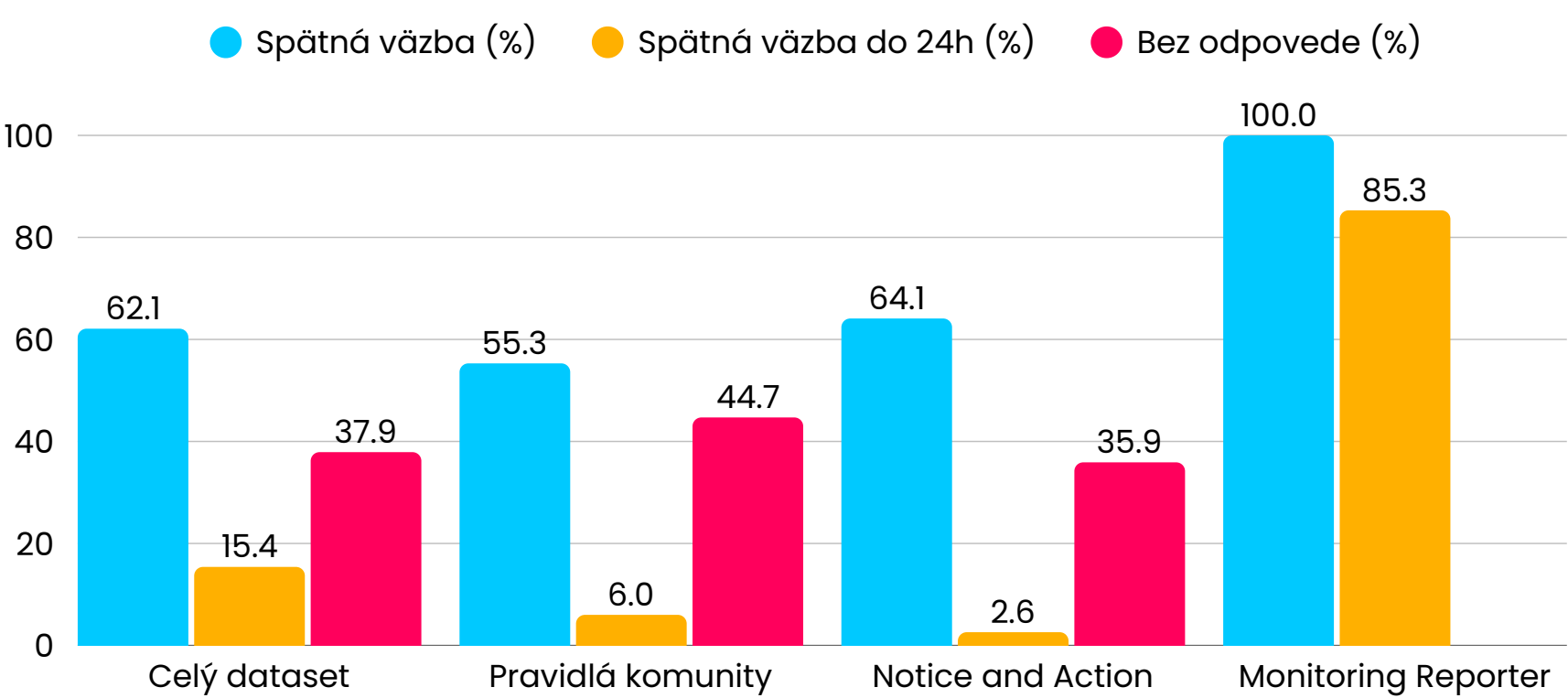
Rýchlosť reakcie na nahlásenia rátame ako čas, ktorý ubehol od nahlásenia do momentu, kedy nám platforma zaslala odpoveď s informáciou, či bol obsah odstránený (alebo inak obmedzený). Analýza pracuje s mediánovými časmi. Prípady, kedy platforma neposkytla na nahlásenie spätnú väzbu, boli zahrnuté ako hodnoty "bez odpovede" a operacionalizované ako  $\infty$ .

Výsledky ukazujú extrémne rozdiely v rýchlosti reakcií medzi platformami aj medzi jednotlivými nahlasovacími kanálmi.



**Graf 5:** Mediánové časy reakcie zo strany platforiem. Os X má na mierke maximum 168 hodín (teda 7 dní). Údaj pre nahlasovanie pod mechanizmom Notice and Action na Facebooku sa do grafu nezmestil – čiara by bola v skutočnosti 6-násobne dlhšia. Naopak krátke reakčné časy TikToku na mierke takmer nevidieť.

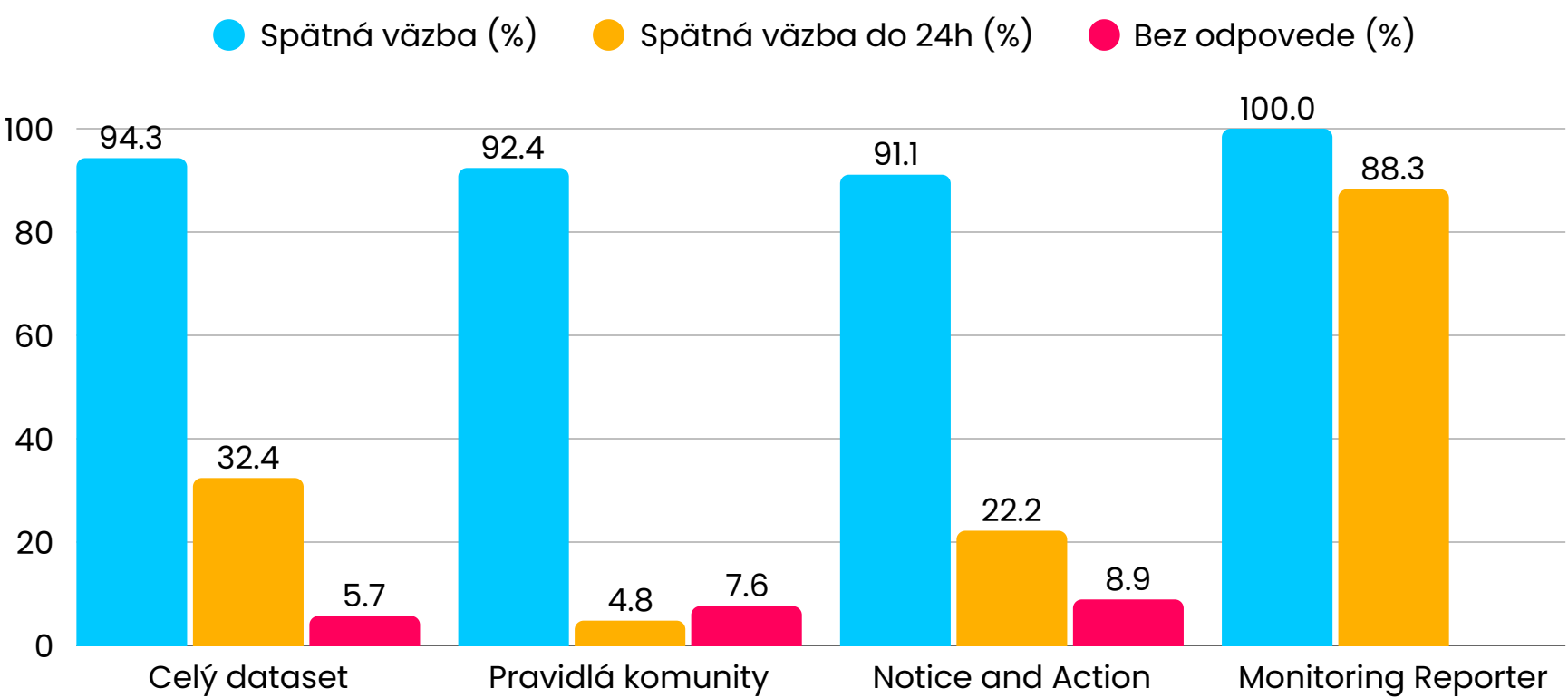
**Facebook** na **37,9 %** nahlásení nereagoval vôbec. Reakcia do 24 hodín – považovaná za optimálnu hodnotu podľa Kódexu správania pre boj proti nezákonným prejavom online – sa vyskytla len v približne **15 %** prípadov, predovšetkým pri špecializovanom kanáli pre Monitoring Reporters, kde sme zaznamenávali vynikajúcu responzivnosť. To nie je prípad štandardných kanálov, kde Facebook vôbec nereagoval na minimálne tretinu nahlásení.



Graf 6: Responzivnosť Facebooku (vyjadrené v %).

Medián pre kanál Monitoring Reporters bol **15,3 hodín**, zatiaľ čo pri nahlásovaní pod mechanizmom Notice and Action dosiahol **41 dní**. Aj po vylúčení prípadov bez odpovede zostáva priemerný čas reakcie (priemer hodnôt časov, kde nastala akákoľvek odpoveď) takmer 23 dní, čo z pohľadu mitigácie rizík robí tento mechanizmus prakticky nefunkčným.

**Instagram** vykazoval výrazne lepšie výsledky. V prípade každého nahlasovacieho kanála spracovala platforma viac ako **90 %** nahlásení. Ťažkopádnejšiu reakciu však sledujeme v prípade nahlasovania pod mechanizmom Notice and Action a najmä pri nahlasovaní porušenia pravidiel komunity, kedy platforma do 24 hodín spracovala menej ako **5 %** nahlásení. V prípade špecializovaného kanála sledujeme podobne ako v prípade Facebooku stopercentnú responzivnosť a prevažne optimálnu rýchlosť reakcie.

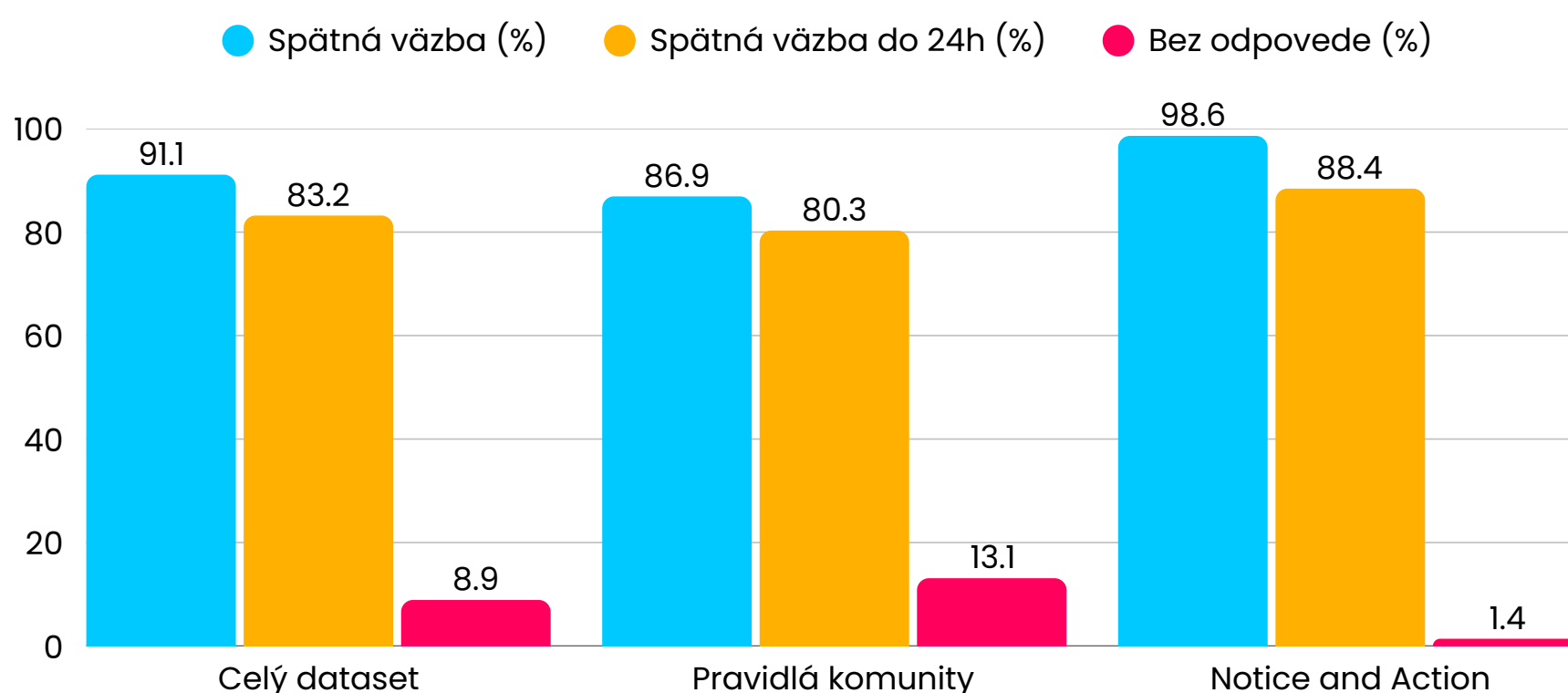


Graf 7: Responzivnosť Instagramu (vyjadrené v %).

Nameraný medián pri nahlasovaní porušenia pravidiel komunity bol takmer týždeň (**6,5 dňa**), čo v kombinácii s oneskoreniami a prípadmi bez odpovede zhoršuje typickú skúsenosť používateľa. V prípade, že rovnaký používateľ nahlásil obsah ako nezákonný pod mechanizmom Notice and Action, sa medián skrátil o takmer 4 dni. Pri špecializovanom kanáli pre Monitoring Reporters Instagram reagoval s mediánom **12,5 hodiny** a maximálnym časom kratším než 48 hodín.

**TikTok** predstavuje kontrastný model. Medián pre celý dataset bol **30 minút**, čo značí vysokú výkonnosť. To, že túto exaktnú hodnotu mala viac než polovica odpovedí, zároveň naznačuje vysoký stupeň automatizácie spracovávania nahlásení.

Veľmi rýchle spracovanie nahlásení, aj vysokú celkovú responzívnosť zaznamenávame na oboch dostupných kanáloch – pri nahlasovaní obsahu ako porušenia pravidiel komunity, aj ako nezákonného obsahu pod mechanizmom Notice and Action. V oboch prípadoch reagoval TikTok na viac ako **80%** nahlásení do 24 hodín. Lepšie výsledky sme zaznamenali v prípade, že sme obsah nahlasovali ako nezákonný, kedy platforma nereagovala len na 1 z nahlásení (**1,4 %**).



**Graf 8:** Responzívnosť TikToku (vyjadrené v %).

Rýchlosť sa však nepretavila do primeranej miery odstránenia obsahu. Vidieť to najmä pri nahlasovaní porušenia pravidiel komunity, kde bola miera odstraňovania extrémne nízka (**3,3 %**).

**YouTube** počas celého sledovaného obdobia nereagoval na žiadne z nahlásení. Aj napriek stretnutiu so zástupcami platformy a ďalším opakovaným pokusom o komunikáciu, zostáva enforcement platformy pri slovenskom obsahu prakticky nefunkčný. To predstavuje významné systémové riziko pre slovenských používateľov.



## Závery a odporúčania

Nenávistné prejavy v slovenskom online priestore nie sú okrajovým javom, ale stabilnou súčasťou digitálneho ekosystému. Naratívne vzorce poukazujú na dlhodobé spoločenské nerovnosti a stigmatizáciu, ktoré sa v online prostredí nielen reprodukujú, ale často aj eskalujú do extrémnych podôb. Závažnosť identifikovaného obsahu – vrátane explicitných výziev na násilie – zároveň ostro kontrastuje s tým, ako nevyrovnane a často nedostatočne naň platformy reagujú. Výsledky monitoringu upozorňujú, že ochrana pred nenávistným obsahom je v praxi skôr nepredvídateľná, poskytovaná nerovnomerne a že dostupnosť ochrany je na viacerých platformách podmienená postavením nahlasovateľa, a nie závažnosťou porušenia alebo mierou rizika, ktorú konkrétny obsah predstavuje.

Napriek tomu, že na Instagrame zaznamenávame pozitívnejšie výsledky ako na platforme Facebook, vykazujú obe sesterské platformy uspokojivé výsledky iba v prípade nahlasovania cez špecializovaný kanál, ktorý je dostupný len úzkemu kruhu expertných organizácií. Instagram však v monitoringu predstavuje platformu, ktorá je pri spracovaní nahlásení relatívne konzistentnejšia. V prípade Facebooku vykazujú kanály určené bežným používateľom extrémne oneskorenia a vysoký podiel prípadov bez odpovede. TikTok sa vyznačuje mimoriadne rýchlym spracovaním nahlásení, pravdepodobne vďaka vysokej miere automatizácie spracovávania nahlásení – to však očividne nevedie k primeranej miere odstraňovania ani pri závažnom obsahu. YouTube sa v sledovanom období ukázal ako platforma, kde mechanizmy nahlasovania systematicky zlyhali, a to aj napriek aktívnej snahe digiQ o diskusiu tohto problému so zástupcami platformy.

Zistenia ukazujú, že rovnaký alebo veľmi podobný obsah je v praxi odstraňovaný podľa toho, kto a akým kanálom ho nahlási. Ak nahlasovanie funguje rýchlo a účinne len v špecializovaných kanáloch, zatiaľ čo bežní používatelia nemajú k dispozícii funkčný mechanizmus, dochádza k narušeniu dôvery v digitálne služby ako bezpečný verejný priestor.

## Odporúčania pre stakeholderov

### Pre online platformy

- **Spriehľadniť účinnosť nahlasovacích kanálov.** Platformy by mali vo svojich správach o transparentnosti uvádzať dezagregované údaje o miere odstránenia a rýchlosti spracovania a spätnej väzby, rozdelené podľa toho, či ide o nahlásenia bežných používateľov alebo subjektov s prioritným statusom. Takéto členenie je nevyhnutné pre objektívne hodnotenie, či mechanizmy nahlasovania fungujú pre všetkých používateľov, nielen pre úzku skupinu aktérov. Nahlasovanie musí zostať funkčným nástrojom ochrany, najmä vzhľadom na to, že zapojenie sa do nenávistnej diskusie s cieľom konfrontovať útočníka, môže kvôli algoritmom kontraproduktívne zvyšovať jeho dosah.

- **Posilniť rýchlosť moderácie obsahu.** Rýchle spracovanie nahlásení je kľúčové pre zmierňovanie rizík nenávistných prejavov, vrátane ich dopadov na mentálne zdravie používateľov a radikalizáciu zraniteľných osôb.
- **Znížiť podiel nahlásení bez odpovede.** Vysoký podiel nahlásení, na ktoré platformy vôbec nereagujú, je zásadným problémom. Platformy by mali upraviť postupy tak, aby sa stav „bez odpovede“ stal výnimočným javom, nie bežnou skúsenosťou používateľa.
- **Vyhnuť sa ďalšiemu znižovaniu počtu moderátorov obsahu,** obzvlášť v jazykoch s nedostatočným pokrytím tak, aby nedochádzalo k výhradnému používaniu strojového spracovania nahlásení. To totiž nemusí dostatočne zohľadňovať kontext nenávistného prejavu.

### Pre národné autority a inštitúcie

- **Presadzovať plnenie povinnosti oznamovania nezákonného obsahu** národným autoritám zo strany platforiem. Dáta ukazujú, že platformy zachytávajú prostredníctvom vlastnej činnosti aj nahlásení obrovské množstvo nezákonného obsahu.
- **Zabezpečiť stabilnú podporu a ochranu organizáciám občianskej spoločnosti,** ktoré disponujú kľúčovými poznatkami o šírení nenávistných prejavov v online priestore, vrátane finančnej a odbornej podpory, a tiež ochrany pred útokmi zameranými na diskreditáciu ich činnosti.
- **Systematicky pracovať s dátami a poznatkami aktérov občianskej spoločnosti,** na identifikáciu nezrovnalostí, nerovného prístupu veľmi veľkých online platforiem a širších trendov, vrátane porovnania skúseností v iných európskych krajinách. Inštitúcie by zároveň mali využívať svoj dosah a vymáhaciu moc tam, kde občianska spoločnosť nemá kapacity na systematickú advokáciu.
- **Posilňovať prevenciu** prostredníctvom vzdelávania, najmä rozvojom digitálnej gramotnosti a zručností digitálneho občianstva vo formálnom aj neformálnom vzdelávaní, čo predstavuje dlhodobý nástroj na prevenciu výskytu a dopadu nenávistných prejavov.



**Zdroje**

**Castaño-Pulgarín, S. A., Suárez-Betancur, N., Tilano Vega, L. M., & Herrera López, H. M.** (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, 58. <https://doi.org/10.1016/j.avb.2021.101608>

**DataReportal.** (2025). Digital 2025: Slovakia. DataReportal. <https://datareportal.com/reports/digital-2025-slovakia>

**digiQ.** (2025a). Správa z prieskumu používania sociálnych médií slovenskou mládežou. Report. <https://digiq.sk/wp-content/uploads/2025/09/Sprava-z-prieskumu-digitalna-verzia.pdf>

**digiQ.** (2025b). Results of Social Media Monitoring in Slovakia 2023-2024. Report. <https://digiq.sk/results-of-social-media-monitoring-in-slovakia-2023-2024/>

**Diko, L., & Kőváry Sólmos, K.** (2025, marec). Odsúdený extrémista Slovákom komunikoval s vrahom zo Zámockej už tri roky pred teroristickým útokom, písali si aj o Teplárni. Investigatívne centrum Jána Kuciaka. <https://www.icjk.sk/383/Odsudeny-extremista-Slovakbro-komunikoval-s-vrahom-zo-Zamockej-uz-tri-roky-pred-teroristickym-utokom-pisali-si-aj-o-Teplarni>

**Európska únia.** (2022). Akt o digitálnych službách, 2022, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065>

**Európska únia.** (2025). Kódex správania pre boj proti nezákonným nenávisným prejavom online+, 2025, <https://digital-strategy.ec.europa.eu/en/library/code-conduct-counteracting-illegal-hate-speech-online>

**Hlavanda Gogová, R.** (2025, február). Tragický útok na gymnáziu: K zlyhaniu došlo v troch kľúčových bodoch. Učitelia upozorňujú na šírenie mizogynie. *Hospodárske noviny*. <https://hnonline.sk/tv/relacie/doslova/96194928-tragicky-utok-na-gymnazium-k-zlyhaniu-doslo-v-troch-klucovych-bodoch-ucitelia-upozoruju-na-siren timerie-mizogynie>

**Lupu, Y., Sear, R., Velásquez, N., Leahy, R., Restrepo, N. J., Goldberg, B., & Johnson, N. F.** (2023). Offline events and online hate. *PloS one*, 18(1). <https://doi.org/10.1371/journal.pone.0278511>

**Rada Európy.** (2015). Všeobecné politické odporúčanie ECRI č. 15 o boji proti nenávisným prejavom. <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech-slov/16808b5b08>



## O digiQ

Občianske združenie Digitálna inteligencia, známe ako digiQ, od roku 2017 aktívne prispieva k budovaniu bezpečnejšieho a zodpovednejšieho online prostredia pre všetkých. Zameriava sa na zvyšovanie povedomia o digitálnom občianstve, etike v IT, problematike nenávistných prejavov a ochrane ľudských práv v online priestore.

Prostredníctvom neformálneho vzdelávania, kampaní a programov podporuje rozvoj digitálnych zručností a kritického myslenia – základných pilierov pre zodpovedné správanie na internete. Činnosť digiQ zahŕňa nielen preventívne aktivity, ako workshopy a prednášky, ale aj intervenciu v oblasti aktuálnych výziev v digitálnom priestore. Ako člen Digitálnej koalície na Slovensku aktívne spolupracuje na formovaní lepšej digitálnej budúcnosti. Na národnej úrovni je digiQ členom Výboru pre predchádzanie a elimináciu rasizmu, xenofóbie, antisemitizmu a ostatných foriem intolerancie SR (VRAX). Na európskej úrovni prispieva k boju proti nenávistným prejavom, ich prevencii, ako aj podpore kritického myslenia online členstvom v sieti INACH, Zero Hate a Facing Facts. Zodpovedný prístup k používaniu internetu a sociálnych sietí podporuje v spolupráci so sieťou INSAFE ako jedna z participujúcich organizácií Centra bezpečného internetu na Slovensku.



[www.digiq.sk](http://www.digiq.sk)



digiQ



@digiqsk



digiQ