tech against terrorism europe

Îmbunătățirea măsurilor luate ca răspuns pentru combaterea terorismului

Un ghid în șase pași pentru aderarea la Regulamentul TCO al UE

Sophia Rothut, Heidi Schulze, Diana Rieger, Catherine Bouko și Brigitte Naderer

Traducere neoficială realizată în scop informativ de Autoritatea Națională pentru Administrare și Reglementare în Comunicații (ANCOM)

Funded by the European Union

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

GHENT UNIVERSITY

Proiectul *Tech Against Terrorism Europe* (TATE)

Ghidul de față face parte din proiectul "Tehnologia împotriva terorismului în Europa" (*Tech Against Terrorism Europe* - TATE). Proiectul TATE este finanțat de Uniunea Europeană (ISF-2021-AG-TCO-101080101). Acest proiect sprijină furnizorii mai mici de servicii de găzduire (HSP) în elaborarea unui cadru de măsuri pentru combaterea terorismului și în realizarea rapoartelor de transparență, în conformitate cu Regulamentul UE privind conținutul online cu caracter terorist (TCO) și Directiva (UE) 2017/541. ▶tate.techagainstterrorism.org

Funded by the European Union

Autorii

Sophia Rothut este cercetător asociat la laboratorul profesorului Rieger (LMU München). Ca parte a proiectului UE TATE, ea lucrează la cerințele pentru combaterea conținutului online cu caracter terorist. Principalele sale teme de cercetare privesc radicalizarea online, încetățenirea ideilor radicale și influențatorii politici/de extremă dreapta.

Heidi Schulze este cercetător asociat la Departamentul de Media și Comunicare de la LMU München. La LMU, ea face parte din Laboratorul de Cercetare al profesorului Rieger și studiază dinamica radicalizării online în cadrul proiectului de cercetare la scară largă MOTRA – Monitorizarea sistematică și platforma de transfer cu privire la radicalizare. În cercetările sale, ea se concentrează pe comunicarea radicală/extremistă (de grup) pe platformele sociale alternative și în comunitățile marginale, precum și pe caracteristicile și publicul site-urilor de știri hiperpartizane.

Diana Rieger este profesor titular la Departamentul de Media și Comunicare al LMU München. Ea realizează cercetări privind radicalizarea online, discursul de incitare la ură și efectele conținutului de divertisment. Ea a publicat, de asemenea, lucrări despre dezvoltarea și evaluarea contramăsurilor împotriva radicalizării.

Catherine Bouko este profesor asociat de comunicare și limbă franceză la Universitatea din Gent (Belgia). Principalele sale teme de cercetare sunt comunicarea politică, extremismul și cetățenia digitală, în special comunicarea bazată pe imagine. Din punct de vedere metodologic, ea folosește analiza discursului (multimodal), analiza cantitativă a conținutului, semiotica și etnografia.

Brigitte Naderer este cercetător asociat post-doctoral la Centrul de Sănătate Publică, Departamentul de Medicină Socială și Preventivă, Unitatea de Cercetare a Sinuciderii și Promovare a Sănătății Mintale de la Universitatea de Medicină din Viena. Anterior, până în martie 2023, a lucrat la Departamentul de Media și Comunicare al LMU München. Cercetările ei se concentrează pe educația media, radicalizarea online și efectele mass-media asupra copiilor și adolescenților.

Mulțumiri: Mulțumim *Tech Against Terrorism* pentru susținerea și contribuția la realizarea acestui ghid. Am dori să-i mulțumim și profesorului Maura Conway, pentru sugestiile sale utile.

Disclaimer: Aceasta este o traducere din limba engleză în limba română menită să faciliteze înțelegerea textului. În caz de discrepanțe între cele două texte, versiunea din limba ▶engleză primează.

Cuprins

A. Introducere

Despre ce este vorba în acest ghid și de ce ar trebui să-l citiți…

Ghidul de față abordează **obligațiile impuse furnizorilor de servicii de găzduire (hosting service providers – HSP) prin Regulamentul european privind conținutul online cu caracter terorist (TCO)** adoptat în aprilie 2021. Acest ghid abordează aspectele pe care HSP trebuie să le ia în considerare pentru a combate diseminarea conținutului online cu caracter terorist și oferă sfaturi practice cu privire la punerea în aplicare a diferitelor măsuri în acest sens.

Ghidul se axează în primul rând pe **cerințele minime pe care trebuie să le îndeplinească HSP pentru a se conforma Regulamentului TCO**. În al doilea rând, oferă **îndrumări și sfaturi practice cu privire la măsurile (pro)active relevante pe care HSP ar trebui să le adopte** pentru a face față cu succes complexității ridicate de aplicarea reglementărilor (ex.: ce să faceți când în căsuța dumneavoastră poștală apare un ordin de eliminare) și pentru a fi pregătiți să combată exploatarea platformelor lor în scop terorist.

Ghidul face parte din proiectul Tech Against Terrorism Europe (TATE), finanțat de Comisia Europeană. Consorțiul TATE este format din șapte parteneri: Dublin City University, Ghent University, JOS Project, LMU Munich, Saher Europe, Swansea University și Tech Against Terrorism. TATE își propune să informeze publicul cu privire la Regulamentul UE privind TCO și să sprijine companiile mici și foarte mici din domeniul tehnologiei în luarea măsurilor pentru implementarea acestuia. Alături de alte resurse, TATE a creat acest ghid pentru a facilita pe cât posibil pentru HSP atingerea așteptărilor legale și practice care îi privesc.

Cui se adresează acest ghid?

Ghidul se adresează furnizorilor de servicii de găzduire (HSP) și angajaților acestora, precum și profesioniștilor IT care doresc să implementeze caracteristici tehnice în arhitecturile platformelor pentru combaterea conținutului terorist. Acest ghid își propune să furnizeze informații cu privire la cerințele minime care trebuie îndeplinite pentru a respecta Regulamentul UE privind TCO.

Am elaborat acest ghid și alte resurse TATE pentru a veni în sprijinul HSP mici și foarte mici. Aceste întreprinderi dispun adesea de resurse limitate pentru a combate conținutul cu caracter terorist de pe platformele lor. **Cu toate acestea, este esențial ca HSP mici și foarte mici să nu neglijeze amenințarea, deoarece platformele mai mici sunt mai susceptibile de a fi exploatate de actorii teroriști** (pentru detalii, a se vedea acest raport Tech Against Terrorism).

Care sunt elementele principale ale Regulamentului UE privind conținutul online cu caracter terorist (TCO)?

Regulamentul european privind conținutul online cu caracter terorist (TCO) a intrat în vigoare în iunie 2022 și **obligă HSP să elimine conținutul sau să blocheze accesul la acesta în termen de o oră de la primirea unui ordin de eliminare din partea unei autorități competente**.

HSP trebuie să coopereze cu autoritățile de aplicare a legii, cum ar fi Europol, precum și cu alte autorități relevante, pentru detectarea și eliminarea conținutului cu caracter terorist care ar putea fi prezent pe platformele lor. Pentru a respecta Regulamentul TCO, **HSP trebuie, de asemenea, să pună în aplicare măsuri eficace și proporționale pentru a preveni reîncărcarea conținutului cu caracter terorist** și mai au anumite obligații despre care veți afla în acest ghid.

Cum este structurat acest ghid?

Regulamentul TCO prevede în prezent ca HSP să satisfacă următoarele cerințe:

1) Elaborarea unor condiții ToS adecvate (▶ capitolul 1),
2) Adoptarea de măsuri specifice pentru identificarea și eliminarea conținutului cu caracter terorist (▶ capitolul 2),
3) Stabilirea unor mecanisme eficiente de moderare (▶ capitolul 3),
4) Stabilirea punctelor de contact și a reprezentanților legali (▶ capitolul 4),
5) Stabilirea mecanismelor de notificare a utilizatorilor și de tratare a plângerilor (▶ capitolul 5),
6) Publicarea rapoartelor de transparență (▶ capitolul 6).

În acest ghid, am structurat explicațiile și recomandările conform celor șase cerințe principale de mai sus. La începutul fiecărui capitol, veți găsi un scurt rezumat al conținutului capitolului și principalele puncte abordate în acesta. Ulterior, vi se vor furniza informații detaliate cu privire la ceea ce Regulamentul TCO impune sau îndeamnă HSP să implementeze, precum și sfaturi practice suplimentare pentru a vă apăra platforma împotriva conținutului cu caracter terorist (și a altor tipuri dăunătoare de conținut).

Ce platforme sunt vizate de Regulamentul TCO?

Regulamentul TCO vizează furnizorii de servicii de găzduire, adică inclusiv orice platformă care permite utilizatorilor să distribuie informații publicului prin intermediul serviciilor sale.

Regulamentul TCO se aplică întreprinderilor **HSP de toate dimensiunile și tuturor HSP care își oferă serviciile în UE**. Acesta se aplică, de asemenea, întreprinderilor HSP situate în afara UE atunci când un astfel de furnizor (1) are un număr semnificativ de utilizatori într-unul sau mai multe state membre UE sau (2) își direcționează activitățile către unul sau mai multe state membre UE.

Când se consideră că un HSP este "expus la conținut cu caracter terorist"?

Regulamentul TCO impune obligații specifice pentru acei HSP care sunt "expuși la conținut cu caracter terorist". În conformitate cu art. 5.4, o astfel de situație se creează atunci când un HSP a fost notificat și a primit două sau mai multe ordine de eliminare definitive în ultimele 12 luni de la autoritatea competentă a statului membru în care HSP își are sediul principal sau reprezentantul legal în UE.

Ce este conținutul cu caracter terorist?

Întrucât Regulamentul TCO se referă la conținutul cu caracter terorist diseminat online, este esențial ca acesta să fie definit. Directiva UE 2017/541 pune bazele Regulamentului TCO prin definirea a ceea ce se înțelege drept conținut cu caracter terorist.

Conținutul este considerat a avea caracter terorist atunci când incită la comiterea de acțiuni sau încurajează intenții în favoarea unei cauze teroriste, contribuind astfel direct sau indirect la amenințări privind comiterea infracțiunilor de terorism.

Amenințarea cu comiterea infracțiunilor de terorism este, de asemenea, considerată conținut cu caracter terorist, la fel ca și furnizarea de informații, sprijinul sau finanțarea unor astfel de acte.

Tipuri de infracțiuni de terorism

Infracțiunile de terorism pot include următoarele (Directiva UE 2017/541, art. 3.1):

- Atacuri asupra vieții sau integrității fizice a unei persoane
- Răpire sau luare de ostatici
- Provocarea de distrugeri masive unor instalații și infrastructuri specifice (ex. instalații guvernamentale/publice, sisteme de transport și informatice) care ar putea pune în pericol vieți umane sau ar putea duce la pierderi economice majore
- Confiscarea aeronavelor, navelor sau a altor mijloace de transport public sau de mărfuri
- Fabricarea, deținerea, achiziționarea, transportul, furnizarea sau utilizarea de explozivi sau arme și cercetarea și dezvoltarea armelor chimice, biologice, radiologice sau nucleare
- Eliberarea de substanțe periculoase sau provocarea de incendii, inundații sau explozii și perturbarea resurselor fundamentale (ex. apă, energie electrică), prin care viața umană este pusă în pericol
- Interferarea sau întreruperea alimentării cu apă, energie electrică sau orice altă resursă naturală fundamentală, prin care viața umană este pusă în pericol.

Conducerea unui grup terorist sau participarea intenționată la activitățile acestuia este, de asemenea, considerată infracțiune (Directiva UE 2017/541, art. 4). Aceasta include furnizarea de resurse (informaționale) (ex. instrucțiuni sau materiale pentru construirea de arme) pentru cauze teroriste sau finanțarea unor astfel de activități.

Prin astfel de infracțiuni, actorii teroriști urmăresc (a) să intimideze grav publicul, (b) să constrângă în mod nejustificat un guvern sau o organizație internațională să (se abțină să) îndeplinească un anumit act sau (c) să destabilizeze grav sau să distrugă structurile fundamentale politice, constituționale, economice sau sociale ale unei țări sau ale unei organizații internaționale [Directiva UE 2017/541, art. 3 alin. (2)].

În concluzie, conținutul este considerat a avea un caracter terorist atunci când permite, sprijină sau facilitează o infracțiune de terorism sau când include o amenințare privind comiterea unor infracțiuni de terorism.

B. Principalele obligații și recomandări privind Regulamentul TCO

În capitolele următoare, veți găsi informații despre șase componente cheie ale Regulamentului TCO și pașii prin care vă puteți întări platforma împotriva amenințării teroriste. Acestea includ următoarele domenii și întrebări cheie.

Capitolul 1: Elaborarea și aplicarea unor Clauze și Condiții ce interzic conținutul cu caracter terorist

1. Ce sunt Clauzele și Condițiile (Terms of Service -ToS)?
2. De ce este necesar să aveți ToS clare și robuste?
3. Îndrumări și sfaturi practice: Cum să redactați ToS robuste?

Capitolul 2 : Măsuri specifice pentru identificarea și eliminarea conținutului (cu caracter terorist)

1. De ce este necesară stabilirea proceselor pentru identificarea conținutului ilegal și dăunător?
2. Procesul pentru identificarea conținutului cu caracter terorist
3. Îndrumări și sfaturi practice: Ce vă ajută să evaluați dacă un anumit conținut este ilegal?
4. Ce ar trebui să faceți dacă vedeți lucrurile diferit? Cum să contestați un ordin de eliminare primit

Capitolul 3: Instituirea unor mecanisme eficace de moderare a conținutului online cu caracter terorist

1. Ce este moderarea conținutului și de ce este necesară în unele cazuri?
2. Îndrumări și sfaturi practice: Cum ar trebui implementată moderarea conținutului?
3. Abordări alternative ale moderării

Capitolul 4: Stabilirea punctelor de contact și a reprezentanților legali

1. Ce sunt punctele de contact și reprezentanții legali?
2. De ce este necesar să avem un punct de contact sau un reprezentant legal?
3. Ce este autoritatea competentă a unui stat membru UE și cum îi pot contacta?

Capitolul 5: Instituirea unui sistem pentru notificarea utilizatorilor și primirea reclamațiilor privind conținutul eliminat

1. De ce este necesară instituirea unui mecanism transparent de tratare a reclamațiilor?
2. Care sunt cerințele pentru un sistem de reclamații?
3. Cum trebuie tratate reclamațiile și care sunt rezultatele posibile?
4. Îndrumări și sfaturi practice: Ce elemente sunt utile în stabilirea unui sistem de reclamații?

Capitolul 6: Sprijin practic și consiliere privind realizarea rapoartelor de transparență

1. Ce sunt rapoartele de transparență?
2. De ce sunt necesare rapoartele de transparență?

3. Procesul de pregătire a rapoartelor de transparență
4. Ce informații și indicatori trebuie incluși în raportul de transparență?

Ghidul explică **aspectele Regulamentului TCO care sunt cele mai relevante pentru platforme, oferind în același timp sfaturi practice cu privire la măsurile proactive care pot fi luate pentru a contracara răspândirea conținutului online dăunător**. Astfel de măsuri sunt esențiale pentru pregătirea platformelor împotriva exploatării serviciilor lor în scopuri teroriste – sau, în limbajul Regulamentului TCO: pentru a putea gestiona situația în momentul în care primul sau orice alt ordin de eliminare ulterior ajunge în căsuța dvs. poștală.

Capitolul 1

Elaborarea și aplicarea unor clauze și condiții ce interzic conținutul cu caracter terorist

Sinteză: Conținutul și principalele puncte ale acestui capitol

- Clauzele și condițiile de utilizare (ToS) sunt un **acord obligatoriu** între utilizator și HSP, care definește utilizarea adecvată și permisă a platformei.
- HSP (1) își stabilesc **strategia de combatere a diseminării conținutului cu caracter terorist** în ToS și (2) **interzic** diseminarea conținutului cu caracter **terorist.**
- De asemenea, în plus față de domeniul de aplicare al Regulamentului TCO, HSP pot și ar trebui să ia în considerare interzicerea mai multor forme de conținut dăunător (ex. conținut extremist, discursuri de incitare la ură).
- ToS sunt o **necesitate legală** în temeiul Regulamentului TCO și pot deveni, de asemenea, **instrumente utile de protecție pentru platforme**.
- Pentru ca ToS să fie cât mai robuste posibil, sunt necesare mai multe elemente, inclusiv o definiție a conținutului cu caracter terorist și comunicarea strategiei HSP de combatere a conținutului cu caracter terorist pe platforma sa.

Clauzele și condițiile **(ToS) adecvate reprezintă baza pentru interzicerea și, în consecință, gestionarea conținutului cu caracter terorist**. Regulamentul TCO îndeamnă în mod explicit HSP să își stabilească "politica lor de prevenire a diseminării conținutului cu caracter terorist, inclusiv, după caz, o explicație pertinentă cu privire la funcționarea măsurilor specifice, inclusiv, după caz, la utilizarea unor instrumente automatizate" [Regulamentul TCO, articolul 7 alineatul (1)].

În plus, depășind domeniul de aplicare al Regulamentului TCO, HSP pot și ar trebui **să ia în considerare interzicerea mai multor forme de conținut dăunător** (de exemplu, conținut extremist, discurs de incitare la ură, promovarea violenței) în ToS. În acest sens, HSP pot contribui la crearea unui cadru pentru o cultură digitală civilă.

1. Ce sunt clauzele și condițiile (ToS)?

ToS sunt reguli stabilite de și pentru platforme specifice, care definesc (1) responsabilitățile HSP față de utilizatorii lor și (2) comportamentul și conținutul adecvat și permis, dar și ce este interzis pe platforma respectivă. Utilizatorii trebuie să accepte aceste condiții dacă doresc să utilizeze serviciile HSP.

Sinonimele utilizate pentru Clauze și Condiții (ToS) sunt, de exemplu, „termeni de utilizare”, „termeni și condiții” sau „standarde comunitare”. Regulamentul TCO folosește expresia „clauze și condiții” și o definește drept „toate clauzele și condițiile, indiferent de denumirea sau forma acestora, care reglementează relația contractuală dintre furnizorul de servicii de găzduire și utilizatorii serviciilor sale” (Regulamentul TCO, Art. 2.8).

2. De ce este necesar să avem ToS clare și robuste?

Vom vedea în amănunt de ce este necesar să avem ToS clare și robuste, din două perspective diferite: cea juridică și cea a unei companii/operațională.

a) Perspectiva juridică

Respectarea legislației (UE)

ToS trebuie să respecte legislația (UE) actuală, inclusiv Regulamentul TCO, dar și Directiva 2017/541 a UE, care consolidează la nivelul UE gestionarea conținutului cu caracter terorist prin (1) furnizarea unei definiții a conținutului cu caracter terorist, (2) stabilirea de sancțiuni pentru acesta (art. 15), (3) consolidarea drepturilor și a sprijinului victimelor și (4) recunoașterea terorismului ca amenințare transfrontalieră transnațională. Atât Regulamentul TCO, cât și Directiva promovează cooperarea la nivelul UE și la nivel internațional. ToS sunt un instrument prin care HSP își pot sublinia angajamentul față de combaterea activității teroriste.

Protecția HSP împotriva răspunderii juridice

ToS sunt destinate să protejeze un HSP de răspundere (atâta timp cât ToS sunt conforme cu legislația aplicabilă). Fiind un contract între HSP și utilizatorii săi, ToS reprezintă un acord obligatoriu din punct de vedere juridic între aceste două părți. ToS permit HSP să stabilească norme pentru utilizarea (in-)acceptabilă în conformitate cu valorile HSP.

Justificarea eliminării (proactive) a conținutului

În plus, ToS clare sunt importante atunci când întâlniți și moderați conținut problematic. HSP pot apela la ToS pentru a justifica moderarea conținutului dacă introduc în acestea informații detaliate privind practicile de conținut interzise, inclusiv interdicții privind terorismul.

b) Perspectiva companiei

Menținerea activității operaționale

În primul rând, este esențial pentru activitatea operațională a HSP și, prin urmare, pentru succesul companiei, ca legislația să fie respectată. HSP au o anumită responsabilitate față de acționarii lor, care se așteaptă ca aceștia să respecte legislația. În plus, nerespectarea regulilor poate atrage răspunderea financiară, iar sancțiunile de reglementare sau vătămarea reputației sunt de asemenea posibile consecințe ale neconformității.

Demonstrarea responsabilității civice a HSP

În al doilea rând, în calitate de actori de marcă ai societății, HSP au responsabilități față de public, atât colectiv, cât și individual. Responsabilitatea față de publicul larg include combaterea activităților ilegale, cum ar fi terorismul și detectarea și prevenirea conținutului cu caracter terorist înainte ca acesta să fie difuzat pe scară largă. Prevăzând în ToS că activitățile ilegale sunt interzise, HSP pun bazele unui mediu online mai sigur și se pot baza pe acestea atunci când acționează împotriva conținutului ilegal.

ToS formulate clar ajută, de asemenea, companiile să își demonstreze responsabilitatea față de membrii individuali ai publicului, construind încredere prin transparență. În acest sens, HSP demonstrează utilizatorilor individuali că vor fi protejați de conținutul dăunător de pe platformă atunci când un astfel de conținut este interzis în mod explicit și implementabil în ToS.

3. Îndrumări și sfaturi practice: Cum să redactați ToS robuste?

Următoarele secțiuni oferă îndrumări cu privire la elementele pe care dumneavoastră, în calitate de HSP, ar trebui să le luați în considerare atunci când elaborați ToS în contextul Regulamentului TCO.

Este important să remarcăm faptul că elaborarea ToS este un proces iterativ: ToS pot și ar trebui să fie adaptate dacă este necesar, de exemplu, ca răspuns la sugestiile utilizatorilor și în pas cu evoluțiile legislative naționale sau internaționale. Regulamentul TCO presupune o astfel de adaptabilitate: în cazul în care sunt expuși la conținut terorist, HSP trebuie să își modifice ToS pentru a menționa acest lucru și pentru a **sublinia acțiunile care vor fi luate pentru a combate utilizarea abuzivă a platformei în scopuri teroriste** (Regulamentul TCO, art. 5.1).

În mod special, vă recomandăm să acordați atenție **clarității și structurii** ToS. Studiile arată că utilizatorii petrec puțin timp citind ToS și nu acordă suficientă atenție informațiilor prezentate; ToS acționează mai degrabă ca o reamintire pentru percepția generală a utilizatorilor despre ceea ce este permis și ceea ce nu este[1,2]. Pentru a ajuta parcurgerea și înțelegerea ToS, este recomandabil să investiți timp în proiectarea vizuală a ToS și să lucrați cu titluri clare sau liste de puncte. De asemenea, poate fi util să le reamintiți periodic utilizatorilor prevederile ToS, iar dacă este detectat un comportament suspect, să transmiteți avertismente cu privire la posibilele consecințe (ex. printr-o fereastră pop-up)[2].

La elaborarea și revizuirea ToS, puteți utiliza următoarea **listă de verificare drept ghid pentru optimizarea conformității cu Regulamentul TCO**.

[1] Obar, J. A., & Oeldorf-Hirsch, A. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society*, *23*(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870
[2] Robinson, E. P., & Zhu, Y. (2020). Beyond "I Agree": Users' Understanding of Web Site Terms of Service. *Social Media + Society*, *6*(1), 1–13. https://doi.org/10.1177/2056305119897321

Definiți conținutul cu caracter terorist

Existența unor definiții funcționale ale terorismului și ale conținutului cu caracter terorist este importantă pentru combaterea acestor tipuri de conținut și comportament. Aceste definiții ar trebui să apară în ToS și să fie folosite în evaluarea conținutului. Definițiile existente prezentate și discutate anterior în acest ghid (►vezi definiția, aici), în special cele ale UE bazate pe Directiva UE 2017/541, pot oferi orientări utile în acest sens.

Prezentați-vă strategia de combatere a conținutului cu caracter terorist

Regulamentul TCO obligă HSP expuși la conținut cu caracter terorist să explice în ToS atât strategia lor de combatere a diseminării unui astfel de conținut, cât și orice mijloace automate utilizate (ex. pentru identificarea conținutului interzis; ►capitolul 2). Regulamentul TCO impune, de asemenea, HSP să pună în aplicare măsuri specifice odată ce platformele sunt expuse la conținut cu caracter terorist (ex. mecanisme pentru raportarea conținutului dăunător de către utilizatori, mecanisme de contestație pentru conținutul eliminat; ►capitolul 4) (Regulamentul TCO, art. 7.1).

Folosiți liste de desemnare și luați în considerare interzicerea clară a altor categorii de conținut dăunător

Tipurile de conținut care pot fi dăunătoare pentru utilizatori, pentru societate și pentru platforma în sine sunt diverse. Astfel, HSP ar trebui să ia în considerare următoarele interdicții suplimentare, abordate și de alte cadre juridice, cum ar fi Actul privind serviciile digitale, și, astfel, să interzică conținutul care include, de exemplu, discursuri de incitare la ură sau incitare la violență. La nivelul actorilor, grupurile teroriste pot fi excluse de la utilizarea platformei. Listele de desemnare (listele oficiale ale grupurilor teroriste publicate de state democratice) sunt o bună resursă pentru astfel de interdicții. În plus, autoritățile competente pot alerta HSP cu privire la conținutul care nu poate să fie clasificat drept terorist, dar care este totuși dăunător; evaluarea conținutului în raport cu ToS și eventuala tratare a conținutului fără caracter terorist, dar totuși dăunător este, în ultimă instanță, responsabilitatea HSP (Regulamentul TCO, 40).

Comunicați ce comportament de utilizare este acceptat

ToS ar trebui, de asemenea, să descrie, prin raportare la valorile și scopurile platformelor, ce comportament de utilizare este nu doar acceptabil, ci chiar salutat și încurajat de către HSP. Nu există un standard universal pentru acest lucru – prevederile variază foarte mult în funcție de funcționalitățile și principiile platformei. Așadar, descrierea utilizării acceptate nu este prevăzută în mod explicit drept o cerință pentru a îndeplini obligațiile Regulamentului TCO.

Oferiți informații despre cum poate fi raportat conținutul interzis

O modalitate de a scoate la iveală conținutul ilegal și care încalcă ToS este prin intermediul sistemelor de raportare puse la dispoziția utilizatorilor. Prin urmare, HSP ar trebui să includă o secțiune în ToS în care să explice procesele prin care poate fi raportat conținutul suspectat de nerespectarea ToS.

Determinați consecințele nerespectării ToS

Pentru a se asigura că orice încălcare a ToS este tratată în mod transparent și pentru a rămâne responsabili față de public, HSP ar trebui să stabilească în mod clar în ToS acțiunile care vor fi întreprinse împotriva utilizatorilor și/sau a conținutului în cazul nerespectării ToS. Conform Regulamentului TCO, un ordin de eliminare emis de o autoritate competentă trebuie să aibă ca efect eliminarea conținutului în termen de o oră de la semnalarea acestuia. Ordinele de eliminare pot fi, de asemenea, puse în aplicare prin blocarea conținutului sau prin blocarea geografică a acestuia în UE.

Publicarea rapoartelor anuale de transparență

Atunci când elaborează și actualizează ToS, HSP ar trebui să se angajeze public că vor publica periodic rapoarte de transparență. Sfaturi practice și cerințele din TCO pentru crearea rapoartelor de transparență sunt prezentate în ▶ capitolul 6.

Capitolul 2

Măsuri specifice pentru identificarea și eliminarea conținutului (cu caracter terorist)

Sinteză: Conținutul și principalele puncte ale acestui capitol

- Capacitatea HSP de a identifica un conținut drept terorist este fundamentală pentru punerea în aplicare a Regulamentului TCO și alte cerințe.
- În multe cazuri, deciziile care trebuie luate nu sunt ușoare și pot fi **foarte delicate și contextualizate.**
- **Drepturile fundamentale, precum libertatea de exprimare, trebuie luate în considerare cu atenție**.
- **Având în vedere Regulamentul TCO, HSP nu trebuie să evalueze legalitatea conținutului alertat de o autoritate competentă.** Cu toate acestea, odată ce a fost emis un ordin de eliminare, atât **HSP, cât și furnizorii de conținut (adică utilizatorii) au dreptul de a-l contesta.**
- Este recomandabil să se stabilească un proces de urmat atunci când se identifică conținut cu caracter terorist. Un astfel de proces va cuprinde mai multe etape.
- Evaluarea conținutului include diverse considerații practice, cum ar fi utilizarea listelor de desemnări sau a bazelor de date cu simboluri și metodele de realizare a exercițiilor de punere în balanță necesare.
- În plus, procedurile de revizuire joacă un rol deosebit de important în cazul special al **ordinelor de eliminare transfrontaliere**, atunci când un HSP nu primește ordinul de eliminare de la autoritatea competentă a statului în care își are sediul principal sau reprezentantul legal. În astfel de cazuri **se aplică proceduri speciale**.

Este de importanță fundamentală ca furnizorii HSP să fie în măsură să evalueze dacă anumite conținuturi sunt ilegale sau au un caracter terorist, pentru a detecta proactiv conținutul cu caracter terorist, precum și a contesta în mod legal ordinele de eliminare atunci când HSP consideră că este posibil ca acestea să fi fost emise în mod eronat. Identificarea corectă a conținutului ca fiind ilegal și terorist este importantă și pentru păstrarea conținutului online legal și fără caracter terorist. Astfel de decizii sunt adesea dificile și este **întotdeauna necesar să se pună în balanță prejudiciul pe care îl poate cauza conținutul față de prejudiciul cauzat de încălcarea dreptului fundamental la libertatea de exprimare**.

1. Stabilirea proceselor de identificare a conținutului ilegal și dăunător

Procesele de identificare cuprind o anumită succesiune de pași pe care companiile sau platformele îi fac pentru a determina măsura în care conținutul încalcă Regulamentul TCO sau prevederi din

clauzele și condițiile de utilizare (ToS) ale platformelor. Acești pași ar trebui să ghideze o evaluare metodică și să permită o concluzie argumentată cu privire la ilegalitatea conținutului.

Astfel de decizii prin punere în balanță sunt **necesare pentru** (a) **a lua măsuri proactive** împotriva conținutului cu caracter terorist și/sau dăunător în alt mod și (b) **a contesta cazurile contencioase de ordine de eliminare** emise de autoritatea competentă în temeiul Regulamentului TCO (mai multe informații privind procesul de contestare vor urma în ▶subcapitolul 4).

Măsurile proactive împotriva conținutului cu caracter terorist și a altor tipuri de conținut dăunător se recomandă din mai multe motive.

1. Prin punerea în aplicare a unor astfel de măsuri, HSP demonstrează că sunt de încredere și că au cele mai bune intenții, față de factorii de decizie politică, autoritățile competente și acționariat.
2. Măsurile proactive consolidează încrederea acționarilor, deoarece delimitează în mod pozitiv compania față de răspunderea legală sau de prejudiciile aduse reputației care ar putea decurge din exploatarea teroristă a platformelor.
3. Întreprinderea de acțiuni insuficiente sau inadecvate împotriva conținutului cu caracter terorist, extremist sau a altor tipuri de conținut dăunător poate avea ca rezultat atragerea pe platformă a utilizatorilor care intenționează să distribuie exact un astfel de conținut. Consecința acestui fapt ar fi o prevalență crescută a conținutului interzis, ceea ce ar necesita investiții mai mari de timp, efort și resurse financiare din partea HSP pentru a contracara fenomenul.

2. Procesul de identificare a conținutului cu caracter terorist

1) Definiția conținutului cu caracter terorist

Orice clasificare a conținutului se face prin referire la o definiție. În ceea ce privește Regulamentul TCO, definiția conținutului cu caracter terorist este cea din Directiva UE 2017/541, astfel cum se menționează în ▶introducere. Aceasta definește conținutul cu caracter terorist drept "material" textual, vizual sau sonor, care:

> "instigă sau solicită unei persoane să comită infracțiuni de terorism sau să contribuie la comiterea acestor infracțiuni, îi solicită să participe la activitățile unui grup terorist sau care glorifică activitățile teroriste, inclusiv prin diseminarea de materiale care prezintă un atac terorist. Definiția ar trebui să includă și materialul care oferă instrucțiuni pentru fabricarea sau folosirea explozibililor, a armelor de foc ori a altor arme sau substanțe nocive ori periculoase, precum și a substanțelor chimice, biologice, radiologice și nucleare (CBRN), sau cu privire la alte metode ori tehnici specifice, inclusiv selecționarea țintelor, cu scopul de a comite infracțiuni de terorism sau de a contribui la comiterea acestora." (Regulamentul TCO, 11)

Definiția nu include "materialul diseminat în scopuri educative, jurnalistice, artistice sau de cercetare sau în scopul sensibilizării opiniei publice împotriva activității teroriste" (Regulamentul TCO, 12), iar drepturile fundamentale, precum libertatea de exprimare, libertatea de informare și libertatea

științelor, trebuie întotdeauna luate în considerare cu atenție. "În plus, exprimarea unor puncte de vedere radicale, polemice sau controversate în cadrul dezbaterii publice privind chestiuni politice sensibile nu ar trebui să fie considerată drept conținut cu caracter terorist." (Regulamentul TCO, 12)

2) Sprijin prin utilizarea instrumentelor automatizate

Poate fi adesea util să folosiți instrumente automatizate pentru identificarea preliminară a conținutului potențial problematic. De exemplu, automatizarea poate fi utilizată pentru a identifica anumite cuvinte cheie asociate cu viziuni extremiste asupra lumii sau elemente vizuale, cum ar fi logo-uri sau simboluri ale organizațiilor teroriste. Există, de asemenea, inițiative de detectare automată a conținutului potențial terorist pe platforme prin referință la o bază de date comună (ex. platforma Terrorist Content Analytics Platform [TCAP] deținută de *Tech Against Terrorism* sau baza de date hash-sharing întreținută de *Global Internet Forum to Counter Terrorism* [GIFCT]).

3) Sprijin prin mecanismele de raportare

În cazul în care un HSP este expus la conținut cu caracter terorist – în mod formal, aceasta înseamnă că autoritatea competentă din țara care găzduiește sediul principal al HSP sau reprezentantul său legal a emis cel puțin două ordine de eliminare obligatorii din punct de vedere juridic, iar HSP a fost informat cu privire la acest lucru – Regulamentul TCO impune în mod explicit luarea de măsuri pentru a preveni diseminarea în continuare a unui astfel de conținut (Regulamentul TCO, Art. 5.2). Se recomandă, luarea de măsuri preventive, pentru a demonstra un comportament proactiv. O astfel de măsură poate fi instituirea unui sistem de raportare care permite utilizatorilor să raporteze către HSP conținutul suspect și interzis. Utilizatorii ar trebui să poată indica o categorie în care s-ar încadra conținutului suspect, iar printre categoriile din care pot alege să se afle și terorismul. Acest lucru permite angajaților care procesează rapoartele să prioritizeze și să proceseze mai rapid conținutul suspectat de răspândirea ideilor teroriste. Această preclasificare poate facilita, de asemenea, pregătirea unor rapoarte ulterioare privind transparența – mai multe detalii sunt disponibile în ▶capitolul 6.

4) Atribuirea către și revizuirea de către moderatorii umani

Semnalarea conținutului prin instrumente automate sau raportarea de către utilizatori individuali este doar un prim pas în atragerea atenției către un conținut suspect și potențial periculos. Ulterior, totuși, revizuirea de către moderatori umani este de obicei necesară înainte de a se adopta o decizie finală privind acțiunea care trebuie întreprinsă. Acești moderatori ar trebui (1) să fie buni cunoscători ai reglementărilor aplicabile platformei, (2) să poată diferenția conținutul interzis de materiale prin care se exercită dreptul la libertatea de exprimare și (3) să aibă cunoștințe amănunțite despre diferitele opțiuni de moderare de pe platformă. Este important ca moderatorii umani să fie instruiți în mod constant cu privire la strategiile online ale teroriștilor, dar este la fel de important, din punct de vedere etic, să fie consiliați cu privire la efectele psihologice ale gestionării conținutului problematic.

5) Decizia privind modul de tratare a conținutului

Procesul descris mai sus culminează cu o decizie privind modul de tratare a conținutului. Este posibil să nu se ia nicio măsură și conținutul să rămână activ dacă, după ce toate implicațiile au fost puse în balanță, conținutul este considerat inofensiv. La cealaltă extremă, în cazul conținutului extrem de problematic, periculos sau interzis în alt mod, cum ar fi apelul la atacuri teroriste, conținutul și utilizatorii pot fi blocați. Există însă și numeroase alte modalități - aflate între aceste extreme - de a trata conținutul - conținutul care nu intră în domeniul de aplicare al Regulamentului TCO poate necesita un tratament mai nuanțat. Detalii și sugestii practice pot fi găsite în ▶capitolul 3 privind strategii eficiente de moderare.

6) Notificarea furnizorului de conținut

În cazul în care conținutul este blocat, furnizorul de conținut trebuie notificat. Pentru aceasta se recomandă un sistem automat de notificare și reclamații. În general, un proces adecvat pentru HSP ar trebui dezvoltat și instituit în acest scop. Detalii și sugestii privind modul în care s-ar putea face acest lucru în conformitate cu Regulamentul TCO sunt disponibile în ▶capitolul 5.

3. Îndrumări și sfaturi practice: Ce este de ajutor în evaluarea caracterului ilegal al conținutului?

În ceea ce privește Regulamentul TCO, HSP nu trebuie să evalueze legalitatea conținutului, această responsabilitate le revine autorităților competente, înainte de a emite un ordin de eliminare. Cu toate acestea, **HSP au obligația de a identifica în mod proactiv conținutul cu caracter terorist de pe platformele lor în momentul în care au fost expuși la astfel de conținut**. Este adesea dificil de evaluat dacă un anumit conținut depășește linia juridică și ar trebui, prin urmare, interzis. În plus, stresul psihologic cauzat de vizionarea conținutului problematic nu trebuie neglijat. Angajații care moderează conținutul ar trebui să aibă regulat oportunitatea de a reflecta cu privire la munca lor și, dacă este necesar, de a primi sprijin psihologic.

Pentru a evalua ilegalitatea conținutului, veți găsi mai jos câteva îndrumări și sfaturi practice. Înainte însă, vom arunca o privire asupra părților relevante despre evaluarea conținutului din Regulamentul TCO:

Recitalul 11 prevede, printre altele:

> "Atunci când evaluează dacă un material reprezintă un conținut cu caracter terorist în înțelesul prezentului regulament, autoritățile competente și furnizorii de servicii de găzduire ar trebui să ia în considerare factori precum natura și modul de formulare a declarațiilor, contextul în care au fost formulate și potențialul acestora de a genera consecințe prejudiciabile pentru securitatea și siguranța persoanelor." (Regulamentul TCO, 11)

În plus, drepturile fundamentale trebuie întotdeauna puse în balanță:

> "Atunci când se stabilește dacă materialul furnizat de un furnizor de conținut constituie „conținut cu caracter terorist" astfel cum este definit în prezentul regulament, ar trebui să se țină seama, în special, de dreptul la libertatea de exprimare și de informare, inclusiv libertatea și pluralismul mass-mediei și de libertatea artelor și a științelor. În special în cazurile în care furnizorul de conținut deține responsabilitate editorială, orice decizie privind eliminarea materialului diseminat ar trebui să țină seama de standardele jurnalistice stabilite de reglementările aplicabile presei sau mass-mediei în conformitate cu dreptul Uniunii, inclusiv cu Carta." (Regulamentul TCO, 12)

În cazul în care HSP primesc un ordin de eliminare, conținutul a fost deja clasificat drept a avea un caracter terorist de către autoritatea competentă. Există însă **două scenarii** în care **este util** ca HSP să aibă competențe pentru a **putea clasifica anumite conținuturi** drept cu sau fără caracter terorist:

1. Dacă un HSP nu a primit încă un ordin oficial de eliminare, dar dorește să **acționeze proactiv** și/sau să impună respectarea propriilor ToS în ceea ce privește conținutul suspect.
2. Dacă un HSP primește un ordin oficial de eliminare, dar **are îndoieli** cu privire la evaluarea conținutului ca fiind cu caracter terorist de către autoritatea competentă și dorește să analizeze ordinul (ca un prim pas către o remediere legală).

Iată mai jos câteva elemente practice care vă pot oferi sprijin în clasificarea conținutului ca având sau nu un caracter terorist.

Utilizați liste de desemnare

Listele de desemnare naționale și internaționale, care numesc organizațiile teroriste, oferă un bun cadru și un punct de referință pentru clasificarea conținutului. De exemplu, adoptarea listelor de desemnare ca bază pentru interzicere sau blocare, la care ar trebui să se facă referire în ToS (▶capitolul 1), oferă sancțiunilor aplicate de platformă sprijinul legii. Recitalul 11 din Regulamentul TCO sugerează în mod explicit că lista Uniunii Europene poate fi utilizată în evaluarea conținutului.

Utilizați cuvinte cheie, baze de date cu simboluri și sigle

Persoanele care examinează manual conținutul ar trebui nu numai să aibă la îndemână definiția conținutului cu caracter terorist, ci și să cunoască temeinic și să își întrețină cunoștințele despre cuvintele-cheie și expresiile utilizate în mod obișnuit de organizațiile teroriste. Cu cât o persoană care evaluează conținutul este mai familiarizată cu un fenomen terorist (ex. de dreapta, de stânga, islamist), cu atât îi este mai ușor să recunoască tacticile de disimulare, cum ar fi așa-numitul "fluier de câini" - adică utilizarea unor cuvinte aparent inofensive care au semnificații ideologice pentru membrii grupărilor respective[3]. O bază de date cu logo-uri și simboluri, adică elemente vizuale, care indică un context terorist, este, de asemenea, indispensabilă pentru evaluarea precisă și rapidă a conținutului.

[3] Åkerlund, M. (2022). Dog whistling far-right code words: the case of 'culture enricher' on the Swedish web. Information, Communication & Society, 25(12), 1808–1825. https://doi.org/10.1080/1369118X.2021.1889639

Includeți factori contextuali

Moderarea ar trebui să ia în considerare contextul în care este probabil să fi fost publicat conținutul. Factorii contextuali relevanți includ, fără a se limita la: (a) condițiile politice, (b) evenimentele curente, inclusiv evoluțiile recente prezentate la știri și (c) circumstanțele culturale care pot modela opiniile cu privire la anumite aspecte. Acest lucru este adesea dificil de judecat – mai ales atunci când utilizatorii sunt anonimi, iar conținutul pe care îl postează conține puține indicii textuale despre identitatea lor sau deloc – dar, în unele cazuri, atunci când în conținut se face referire la factori externi, considerentele de context se dovedesc a fi utile.

În plus, atunci când aveți de-a face cu un **conținut fără caracter terorist, dar dăunător în alt mod** (ex. diferite forme de discurs de incitare la ură), pot fi luate în considerare următoarele aspecte.

Puneți în balanță amploarea daunelor potențiale cauzate de conținut

Conținutul terorist, extremist și alte tipuri de conținut dăunător, cum ar fi discursul de incitare la ură sau la violență, în special atunci când incită la comiterea de infracțiuni de terorism, poate provoca daune semnificative. Atunci când se evaluează conținutul, poate fi utilă analizarea măsurii în care conținutul poate fi un factor cauzator de astfel de daune. Cu cât sunt mai mari daunele care pot apărea, cu atât mai repede și mai hotărât trebuie acționat. Printre daune ar trebui să se numere și efectele psihologice negative asupra celor afectați de conținutul respectiv.

Luați în considerare potențialul impact al eliminării conținutului

Regulamentul TCO solicită asigurarea unui echilibru sensibil între drepturile fundamentale consacrate în Cartă (ex. libertatea de exprimare și de informare) atunci când se are în vedere eliminarea conținutului. În cazul în care eliminarea nu a fost dispusă de o autoritate competentă, există alte modalități prin care poate fi tratat conținutul dăunător – exemple în acest sens pot fi găsite în ▶ capitolul 3.

4. Ce ar trebui să faceți dacă vedeți lucrurile diferit? Cum contestați un ordin de eliminare

Ce e de făcut atunci când rezultatele evaluării HSP și ale autorității competente nu coincid?

Atât HSP, cât și furnizorul de conținut au dreptul de a contesta ordinele de eliminare în temeiul Articolului 9 din Regulamentul TCO. Acest drept reprezintă un important sistem de control și echilibru capabil să susțină drepturile fundamentale.

Contestarea ordinelor de eliminare se face în fața instanțelor din statul membru al UE a cărui autoritate competentă a dispus eliminarea. **HSP sunt obligați să păstreze în condiții de siguranță și pentru o perioadă de șase luni tot conținutul care a fost eliminat (fie ca urmare a unui ordin de eliminare, fie ca urmare a altor măsuri), precum și orice date asociate conținutului (de exemplu, ora publicării, informații despre cont)**. Acest lucru sprijină investigarea și prevenirea infracțiunilor

sau amenințărilor teroriste și asigură faptul că există suficient timp pentru inițierea contestațiilor legale împotriva eliminării conținutului și, dacă este necesar, pentru restabilirea conținutul (Regulamentul TCO, 27 și 28).

Caz special: Ordinele de eliminare transfrontaliere (Regulamentul TCO, art. 4)

În cazul în care HSP primesc un ordin de eliminare din partea unei autorități competente dintr-un alt stat membru al UE decât "autoritatea din țara sa de origine" (ex. autoritatea competentă a statului membru în care HSP își are sediul principal sau reprezentantul legal; detalii privind reprezentanții legali sunt disponibile în ▶ capitolul 4), **se aplică o procedură specială**.

Când emite ordinul de eliminare, autoritatea competentă trebuie să transmită și o copie a acestuia "autorității din țara de origine", care dispune de un termen de 72 de ore pentru a examina ordinul. În cazul în care se constată că ordinul de eliminare încalcă drepturile și libertățile fundamentale consacrate în Cartă, aceasta poate emite o decizie motivată, care poate constitui o obiecție.

HSP trebuie să elimine și să securizeze conținutul la primirea ordinului de eliminare (ca în cazul unui ordin de eliminare emis de "autoritatea din țara de origine"). HSP (și furnizorul conținutului eliminat) pot depune o cerere de reexaminare la "autoritatea de origine" în termen de 48 de ore, care are dreptul să revizuiască ordinul de eliminare și să răspundă în termen de 72 de ore de la primirea cererii, după ce a notificat autoritatea competentă care a dispus inițial eliminarea că se efectuează o reexaminare.

După ce autoritatea care a efectuat reexaminarea a emis o decizie motivată, actorii relevanți (și anume autoritatea inițială, HSP, furnizorul de conținut și Europol, dacă este cazul) sunt notificați și, dacă a fost detectată o încălcare, HSP poate restabili imediat conținutul.

Capitolul 3

Stabilirea unor mecanisme eficace de moderare a conținutului online cu caracter terorist

Sinteză: Conținutul și principalele puncte ale acestui capitol

- În moderarea conținutului, **conținutul utilizatorului este analizat pentru a evalua dacă au fost respectate sau încălcate prevederile legale** (ex. Regulamentul TCO, DSA) **și regulile specifice ale platformelor** (ex. ToS).
- **Dacă una din reguli a fost încălcată, conținutul va fi 'moderat'.** Cu alte cuvinte, sunt luate măsuri pentru a limita răspândirea conținutului.
- Atunci când se primește un **ordin de eliminare** de la o autoritate competentă în baza Regulamentului TCO, **moderarea conținutului include întotdeauna eliminarea** respectivului conținut cu caracter terorist, deși HSP și utilizatorii pot contesta această acțiune dacă nu sunt de acord cu ea.
- Mai multe elemente trebuie avute în vedere înainte, în timpul sau imediat după întreprinderea de către HSP a măsurilor de moderare, precum și ulterior. Acestea includ notificarea utilizatorilor al căror conținut a fost moderat și opțiunea de a contesta decizia.
- Dacă platformele iau **în mod proactiv acțiuni contra altor forme de conținut potențial dăunătoare**, precum discursul care incită la ură sau insultător, dincolo de Regulamentul TCO, pot fi luate în considerare alte **abordări alternative de moderare**.

1. Ce este moderarea conținutului și de ce este necesară în unele cazuri?

Moderarea conținutului implică **analiza conținutului generat de utilizator** (UGC) pe internet pentru a evalua **gradul de adecvare** al conținutului prin raportare la regulile platformelor (ex. ToS) și la cadrul de reglementare (ex. Regulamentul TCO, DSA)[4]. **Regulile platformelor și cerințele legale sunt implementate prin moderarea conținutului** astfel ca acțiunea necesară să fie luată contra conținutului interzis și/sau problematic. HSP trebuie să fie **transparenți** cu privire la măsurile specifice luate pentru identificarea și eliminarea conținutului cu caracter terorist, inclusiv cu privire la procesele proactive și reactive de moderare a conținutului și orice instrumente automate utilizate.

Moderarea conținutului este o procedură foarte importantă, dar și foarte complicată și **sensibilă**. Deținerea de resurse tehnice și expertiză adecvate poate fi o povară financiară semnificativă pentru

[4] Roberts, S.T. (2017). Content Moderation. In L. Schintler & C. McNeely (eds.): *Encyclopedia of Big Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_44-1

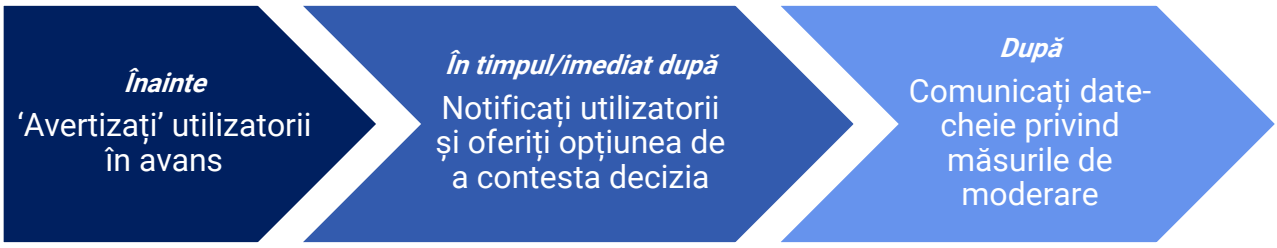
companiile mai mici. Mai mult, conținutul cu caracter terorist poate fi produs în toate limbile și sunt necesare foarte multe resurse pentru a asigura că moderatorii pot evalua conținutul în toate aceste limbi.

Drepturile fundamentale consacrate în Declarația ONU privind Drepturile Omului, în special dreptul libertății de exprimare, trebuie mereu cântărite față de răul cauzat de conținutul în cauză. Pe de altă parte, HSP au responsabilitatea socială de a limita conținutul dăunător de pe platformele lor.

O moderare a conținutului comprehensibilă și transparentă contribuie la protejarea utilizatorilor și a *stakeholder*-ilor față de efectele conținutului dăunător și întărește încrederea. Din perspectiva HSP, moderarea conținutului poate susține limitările de răspundere, respectarea legilor aplicabile și protecția contra vătămării reputaționale cauzate de utilizarea defectuoasă a platformei.

Moderarea conținutului nu înseamnă în mod necesar eliminarea conținutului. În vreme ce ștergerea conținutului poate fi obligatorie ca urmare a cerințelor legale - cum este cazul ordinelor de eliminare bazate pe Regulamentul TCO – există și alte căi de a aborda conținutul dăunător. În funcție de cât de problematic și grav este conținutul și de cât de dăunător poate fi, pot fi luate în considerare opțiuni alternative. Exemple și sugestii de strategii alternative de moderare sunt descrise în ▶ secțiunea 3 din acest capitol.

Stilurile de moderare variază de la o platformă la alta, în funcție de funcționalitatea acestora, serviciile oferite, valorile platformei și toleranța la risc[5]. Este important să fiți cât mai transparent posibil cu moderarea conținutului. Aceasta include diverși factori care trebuie luați în considerare înainte, în timpul și după ce are loc moderarea pentru a vă asigura că utilizatorii cunosc acțiunile de moderare pe care le-ar putea suporta conținutul lor.



1. **Înainte:** 'Avertizați' utilizatorii în avans

Informați-i pe utilizatori într-un stadiu incipient ce acțiuni de moderare a conținutului ar putea fi luate pentru un comportament sau un conținut interzis prin ToS sau prin lege. Aceasta înseamnă: să includeți informații privind strategiile dvs. de moderare în acordul obligatoriu cu utilizatorii, adică în

[5] Roberts, S.T. (2017). Content Moderation. In L. Schintler & C. McNeely (eds.): *Encyclopedia of Big Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_44-1

ToS (►capitolul 1). Acest lucru vă ajută, de asemenea, să vă asigurați că începeți să construiți o relație de încredere cu persoanele înainte sau în momentul intrării în relația cu utilizatorul și că îl asigurați într-un mod obligatoriu și lipsit de ambiguitate de faptul că nu veți tolera pe platforma dvs. un conținut problematic și ilegal, și că dvs., în calitate de HSP, doriți să protejați utilizatorii de un astfel de conținut.

2. **În timpul (sau imediat după acțiune):** Notificați utilizatorii al căror conținut a fost blocat sau moderat și oferiți-le opțiunea de a contesta decizia

Articolul 11 din Regulamentul TCO obligă HSP, în primul rând, să informeze utilizatorul care a furnizat (adică a creat și / sau încărcat) conținutul interzis (terorist) că a fost blocat de platformă și, în al doilea rând, să furnizeze utilizatorului informații asupra motivelor acțiunii sau să furnizeze ordinul de eliminare la cererea utilizatorului. Notificarea poate fi ascunsă temporar furnizorilor de conținut pentru o perioadă de cel mult șase săptămâni, în cazul în care autoritatea competentă consideră că această comunicare a eliminării către furnizorul de conținut prezintă un pericol deosebit. Atunci când nedivulgarea rămâne importantă și adecvată, autoritatea competentă poate prelungi perioada cu încă șase săptămâni (Regulamentul TCO, Art. 11.3). Regulamentul TCO (Art. 10) prevede mecanisme specifice pentru gestionarea unor astfel de situații. Este recomandabil să se instituie un proces automat standardizat prin care furnizorii de conținut să fie notificați cu privire la ștergeri și să li se ofere posibilitatea de a depune o plângere împotriva unor astfel de acțiuni. Acest mecanism de notificare și de contestare ar trebui, de asemenea, să fie utilizat în cazurile care nu intră sub incidența Regulamentului TCO, dar care au avut loc prin măsuri (pro)active din partea HSP.

3. **După:** Comunicați public cu regularitate date-cheie privind măsurile de moderare a conținutului

Transparența cu privire la deciziile și rezultatele moderării conținutului este importantă pentru a stabili încrederea și responsabilitatea între HSP și utilizatorii lor și este impusă din ce în ce mai mult de diferite tipuri de reglementări online. Rapoartele de transparență captează informații privind deciziile de moderare ale platformelor, incluzând nu numai numărul și tipul încălcărilor identificate, ci și modul exact în care au fost tratate aceste încălcări. (Mai multe detalii privind rapoartele de transparență și cerințele de transparență din TCO pot fi găsite în ►capitolul 6.)

2. Îndrumări și sfaturi practice: Cum ar trebui implementată moderarea conținutului?

Este important ca HSP să aibă un **proces clar de moderare a conținutului**, care poate fi **personalizat** în funcție de serviciile și nevoile specifice ale activității. Este esențial să se identifice mai întâi conținutul interzis și să se poată evalua ilegalitatea acestuia, astfel cum se arată mai detaliat în ►capitolul 2. Vă reamintim pașii cheie:

1. Definiția conținutului terorist
2. Sprijin prin utilizarea instrumentelor automate
3. Sprijin prin mecanismele de raportare

4. Atribuirea către și revizuirea de către moderatori umani
5. **Decizia privind modul de tratare a conținutului (adică decizia privind moderarea conținutului)**
6. Notificarea furnizorului de conținut

În această secțiune ne vom concentra pe *pasul 5*. În ▶capitolul 2, am denumit acest pas "Decizia privind modul de tratare a conținutului". După ce au primit un ordin de eliminare din partea unei autorități competente, HSP trebuie să elimine conținutul și să aibă posibilitatea de a contesta ordinul. De asemenea, HSP pot decide să identifice și să elimine în mod proactiv conținutul cu caracter terorist sau interzis (adică în absența ordinelor de eliminare).

Acest ghid distinge cinci tipuri diferite de moderare: pre-, post-, reactiv, distribuit și automatizat. Explicații scurte ale fiecăruia, precum și avantajele și dezavantajele acestora sunt redate mai jos.

Tipul de moderare	Explicație
Pre-moderare	Moderatorii examinează conținutul înainte de publicare. **Pro:** Grad ridicat de siguranță și de respectare a standardelor legale și a regulilor specifice ale platformei în privința conținutului respectiv **Contra:** Efort foarte mare, personal și (eventual) costuri financiare
Post-moderare	Moderatorii revizuiesc conținutul imediat după publicare. **Pro:** Permite interacțiunea rapidă a utilizatorilor cu conținutul (datorită publicării imediate) **Contra:** Cantitate semnificativă de timp, personal (și eventual) costuri financiare; În cazul conținutului interzis/dăunător, utilizatorii sunt expuși la conținutul respectiv
Moderare reactivă	Moderatorii examinează conținutul după ce utilizatorii l-au raportat. **Pro:** Mai puțin consumator de resurse pentru HSP; Consolidează relația de încredere cu utilizatorii (prin posibilitatea raportării) **Contra:** Responsabilitatea utilizatorilor; în cazul conținutului interzis/dăunător, utilizatorii sunt încă expuși conținutului; potențial pentru alarme false care necesită eforturi suplimentare din partea HSP
Moderare distribuită	Comunitatea de utilizatori acționează ca moderatori, adesea printr-un mecanism de vot pozitiv și negativ care măsoară credibilitatea conținutului și este utilizat în cele din urmă pentru a determina acoperirea publică a conținutului (conținutul cu multe voturi pozitive se clasează mai bine și obține o acoperire mai mare). **Pro:** Mai puțin consumator de resurse din partea HSP; Încurajează implicarea utilizatorilor; Autoreglementare **Contra:** Responsabilitatea utilizatorilor; în cazul conținutului interzis/dăunător, utilizatorii sunt expuși conținutului; Susceptibilitatea la un comportament manipulator coordonat din partea utilizatorilor rău intenționați
Moderare automată	Modele bazate pe inteligența artificială (de exemplu, filtre, algoritmi) funcționează ca moderatori. **Pro:** : Consumă mai puține resurse odată ce este configurat; Detectarea timpurie, rapidă și extrem de scalabilă a conținutului potențial malițios

Tipul de moderare	Explicație
	Contra: Susceptibilitatea la erori, în special în ceea ce privește ștergerea eronată a conținutului (problematică în contextul libertății de exprimare, deci e importantă evaluarea umană); necesită întreținere continuă și adaptare la noile evoluții

(Bazat pe Grimes-Viort, 2010)[6]

Se recomandă combinarea mai multor abordări ale moderării conținutului. Moderarea reactivă, de exemplu, poate fi ușor combinată cu pre- sau post-moderarea. În cele mai multe cazuri, **combinațiile sunt foarte utile și, în unele cazuri, pot fi chiar necesare**. Deciziile automate de moderare trebuie să fie întotdeauna revizuite sau cel puțin informate prin examinare umană pentru a evita încălcările sistematice ale drepturilor fundamentale ale utilizatorilor și riscurile la adresa libertății de exprimare.

Rețineți că actorii extremiști și teroriști pot folosi **tactici pentru a evita metodele cunoscute de moderare a conținutului**, pentru a opera sub radar și, astfel, pentru a eluda moderarea, în special mecanismele automate de moderare a conținutului. Tacticile populare includ utilizarea *scurtării* adreselor URL pentru a evita filtrele sau blocarea site-urilor web, *oglindirea* contului și a conținutului, care implică postarea sau crearea de conținut / conturi identice de mai multe ori pentru a copleși moderatorii și pentru a avea copii de rezervă disponibile în caz de ștergere sau *alegerea deliberată a ortografiei (incorecte) a unui cuvânt* pentru a trece de filtrele automate de cuvinte. Gama de tehnici evazive subliniază importanța moderării umane.

Implicarea dvs. în acest ghid sau participarea dvs. (certificată) la cursul online, care este, de asemenea, oferit ca parte a proiectului TATE, arată că sunteți interesat de aceste tendințe. Puteți găsi mai multe exemple de evitare a moderării conținutului, precum și răspunsuri fezabile și eficiente în platforma de partajare a cunoștințelor Tech Against Terrorism, care este disponibilă aici.

3. Abordări alternative de moderare

Moderarea nu trebuie să însemne întotdeauna ștergerea conținutului. Abordările alternative, de tipul celor sugerate mai jos, pot prezenta interes pentru HSP care doresc să modereze conținutul (pro)activ dincolo de cerințele Regulamentului TCO. Este important să subliniem că astfel de abordări alternative de moderare **nu intră în domeniul de aplicare al Regulamentului TCO și pot fi utilizate numai atunci când platformele nu au primit un ordin de eliminare, însă doresc să modereze proactiv conținutul neterorist, dar dăunător**. Când primiți un ordin de eliminare, procedura este clară: trebuie să eliminați conținutul și nicio abordare alternativă de moderare nu intră în discuție.

[6] Grimes-Viort, B. (2010, December 7). 6 types of content moderation you need to know about. *Social Media Today.* https://www.socialmediatoday.com/content/6-types-content-moderation-you-need-know-about

DORIȚI SĂ FACEȚI MAI MULT PENTRU A VĂ MENȚINE PLATFORMA ÎN SIGURANȚĂ?

Ascundeți conținutul

HSP pot ascunde parțial sau complet conținutul și, prin urmare, pot evita blocarea acestuia, în cazul în care consideră că utilizatorii pot găsi conținutul ofensator sau inacceptabil, dar acesta este totuși legitim, legal și permis pe întreaga platformă. Ascunderea conținutului de persoane dintr-un grup vulnerabil sau de persoane aflate într-o țară în care conținutul este ilegal (în timp ce este permis în alte țări) este un astfel de răspuns. Diverse funcționalități tehnice pot fi utilizate pentru a ascunde conținutul, cum ar fi filtrele de conectare sau paywall, un mod securizat (căutare) pentru a afișa conținut adecvat vârstei și blocarea geografică sau temporală.

Decuplați conținutul de mecanismele de recompensare ("dezimplicare")

Dezangajarea privează anumite conținuturi sau anumiți utilizatori de indicatori privind interacțiunile, descurajează activitatea în jurul postării și poate face conținutul în general nesatisfăcător pentru postare. Cu toate acestea, conținutul și contul de utilizator rămân pe platformă. Dezangajarea restricționează vizibilitatea postărilor sau conturilor de pe platformă. Tacticile tipice de dezangajare includ dezactivarea caracteristicilor platformei, cum ar fi, în cazul multor rețele sociale, capacitatea de a aprecia, comenta sau distribui postări, astfel încât postarea să poată fi citită, demonetizarea (adică privarea conturilor de capacitatea de a câștiga bani din conținutul lor) sau eliminarea verificării (adică eliminarea oricărei certificări a identității contului sau a utilizatorului). Astfel de sancțiuni pot implica și pot fi agravate de o schimbare a tratamentului conținutului sau al contului de către algoritmul platformei: retrogradarea conținutului înseamnă că este mai dificil să fie distribuit și promovat pe scară largă prin mecanismele platformei (de obicei, algoritmi de recomandare).

Atașați note pedagogice la conținut

Scopul tacticilor pedagogice sau bazate pe comunicare este de a oferi utilizatorilor informații suplimentare, astfel încât aceștia să poată decide singuri dacă doresc să vadă conținutul sau nu. În cele din urmă, platforma decide ce conținut este furnizat cu astfel de note, ce implică aceste note, despre ce categorie de vătămare sunt avertizați utilizatorii și cât de multe informații suplimentare sunt oferite. O practică binecunoscută utilizată de Twitter (anterior) este de a avertiza utilizatorii că ar putea exista conținut dăunător într-o postare, cum ar fi dezinformarea sau narațiunile conspiraționiste, și de a le permite utilizatorilor să vadă conținutul numai după ce au confirmat în mod activ că doresc să facă acest lucru apăsând pe un buton. În special în cazul conținutului politico-ideologic care ar putea promova radicalizarea,

contradiscursurile și linkurile către informații educaționale pot sensibiliza oamenii cu privire la posibilele efecte ale unui astfel de conținut.

Oferiți responsabilitate utilizatorilor ("abilitarea comunității")
Premisa mecanismelor de moderare bazate pe abilitarea comunității este de a permite utilizatorilor înșiși să creeze spațiul digital pe care și-l imaginează. Astfel de strategii pot prezenta un interes deosebit pentru platformele pentru care ideea de comunitate este importantă sau ale căror practici de moderare depind deja, într-o anumită măsură, de sprijinul utilizatorilor. Aceste abordări de moderare urmează tipul de moderare distribuită. În plus față de funcționalitatea de vot pozitiv și negativ deja menționată, aceasta include și blocarea sau dezactivarea individuală a anumitor conturi, pe care un număr mare de platforme le oferă deja, sau utilizarea administratorilor sau moderatorilor din comunitate. Strâns legat de acesta este conceptul de "notificatori de încredere", care este examinat în Actul privind serviciile digitale (articolul 22). Conceptul se referă la utilizatorii care sunt demni de încredere și competenți să evalueze ilegalitatea conținutului și să îl raporteze (obiectiv și rapid) și care reprezintă interese colective (orientate spre bunăstarea publică), indiferent de platforma online specifică. Conținutul raportat în acest mod ar trebui să fie procesat cu prioritate și rapid.

Aceste abordări alternative de moderare pot fi relevante chiar și în cazul în care platformele nu sunt obligate de Regulamentul TCO să ia măsuri. Indiferent de forma pe care o ia, Regulamentul TCO permite o abordare proactivă: dacă, în cursul măsurilor (pro)active, proprii de moderare, HSP întâlnește conținut care se referă la o amenințare iminentă la adresa vieții sau la un act terorist, acesta trebuie șters, iar autoritatea competentă a statului membru al UE afectat de acesta trebuie informată imediat (Regulamentul TCO, Art. 14.5).

Mai multe detalii despre metodele (tehnice) necesare acestor abordări alternative, precum și avantajele și dezavantajele acestora și studiile de caz, sunt furnizate de Tech Against Terrorism aici.

Capitolul 4

Stabilirea punctelor de contact și a reprezentanților legali

Sinteză: Conținutul și principalele puncte ale acestui capitol

- Regulamentul TCO face distincție între punctele de contact și reprezentanții legali.
- **Punctele de contact trebuie să fie stabilite de fiecare HSP și sunt responsabile** pentru primirea **ordinelor de eliminare** și prelucrarea promptă a acestora.
- **În cazul în care HSP nu sunt stabiliți în UE, trebuie, de asemenea, numit un reprezentant legal.** Această persoană este responsabilă pentru primirea, respectarea și aplicarea Regulamentului TCO. Reprezentantul legal poate, dar nu este obligat, să acționeze și ca punct de contact.
- Fiecare HSP, indiferent dacă a fost sau nu expus la conținut cu caracter terorist, trebuie să stabilească un punct de contact și, dacă este necesar, un reprezentant legal, în conformitate cu Regulamentul TCO.

Regulamentul TCO prevede că *toți* HSP care sunt afectați de acest regulament (intră în domeniul de aplicare al regulamentului prin definiție; ►a se vedea secțiunea privind platformele afectate din introducere) sunt obligate să numească un punct de contact sau un reprezentant legal.

1. Ce sunt punctele de contact și reprezentanții legali?

Punct de contact (Regulamentul TCO, 42 & Art. 15)

- **Scop**: Punctele de contact ale HSP facilitează procesarea imediată a ordinelor de eliminare. Prin urmare, punctul de contact servește numai unor scopuri operaționale.
- **Logistică**: Punctul de contact ar trebui să poată primi și transmite ordinele de eliminare pe cale electronică, indiferent dacă este intern sau externalizat.
- **Resursele necesare**: Punctul de contact trebuie să aibă suficiente capacități tehnice, acces și să dispună de personal astfel încât ordinele de eliminare să poată fi procesate fără întârziere. Întrucât conținutul cu caracter terorist trebuie eliminat în termen de o oră de la primirea ordinului de eliminare, punctul de contact trebuie să fie disponibil 24 de ore din 24, 7 zile din 7.
- **Locație**: Punctul de contact nu trebuie neapărat să fie situat în UE.
- **Comunicare**: Limba în care se poate comunica cu punctul de contact trebuie indicată în informațiile despre disponibilitatea acestuia. Pentru a permite comunicarea între HSP și autoritățile competente ale statelor membre, ar trebui utilizată cel puțin o limbă oficială a UE. Aceasta ar trebui să fie o limbă în care să fie disponibili și ToS ai platformei.

Reprezentant legal (Regulamentul TCO, Art. 17)

- **Necesitate**: În cazul în care HSP nu își are sediul principal în UE, trebuie desemnat un reprezentant legal. Aceasta este o persoană fizică sau juridică aflată într-unul dintre statele membre ale UE în care HSP își oferă serviciile.
- **Scop**: Acest reprezentant legal este responsabil pentru primirea, respectarea și aplicarea Regulamentului TCO și, în special, a ordinelor de eliminare.
- **Resurse**: HSP trebuie să ofere reprezentanților legali puterea, capacitățile și resursele necesare pentru a se conforma Regulamentului TCO și pentru a coopera cu autoritățile competente.
- **Răspundere**: Reprezentanții legali pot fi trași la răspundere pentru încălcarea Regulamentului TCO.
- **Relația cu punctul de contact**: Reprezentantul legal poate acționa și ca punct de contact în același timp, dar nu este obligat să facă acest lucru.

2. De ce este necesar un punct de contact sau un reprezentant legal?

Celelalte capitole prezintă multe opțiuni care pot fi atrăgătoare pentru HSP din punct de vedere comercial, dar principalul motiv și argument pentru **stabilirea unui punct de contact sau desemnarea unui reprezentant legal** se datorează faptului că **este obligatoriu**. Acest lucru poate implica o anumită realocare a resurselor, dar este, fără îndoială, în interesul HSP să respecte legea și să evite consecințele negative ale neconformității - cum ar fi pierderea reputației, a încrederii și a banilor, în cazul în care se impun amenzi - prin desemnarea promptă.

HSP trebuie să permită autorităților competente să revizuiască informațiile privind punctul de contact și, ulterior, să furnizeze notificări electronice (să transmită un ordin de eliminare pe cale electronică). De obicei, acest lucru implică furnizarea unei adrese de e-mail (de exemplu, în secțiunea "Contactați-ne" de pe site-ul HSP) prin care autoritatea competentă poate contacta punctul de contact.

Există două aspecte ale numirii unui reprezentant legal care trebuie subliniate. În primul rând, **identitatea reprezentantului legal trebuie să fie făcută publică** (a se vedea punctul de contact). În al doilea rând, **HSP trebuie să notifice în mod activ numirea "autorității din statul de origine"** (autoritatea competentă a statului membru în care este stabilit reprezentantul legal).

Europol a dezvoltat, de asemenea, o platformă, Plateforme Européenne de Retraits de Contenus illicites sur Internet (Platforma europeană pentru eliminarea conținutului ilicit online) sau **PERCI**, **pentru a sprijini punerea în aplicare a Regulamentului TCO**. Scopul PERCI este de a se asigura că HSP pot primi ordine de eliminare din partea statelor membre printr-un canal securizat comun, în loc de 27 de sisteme separate pentru fiecare stat membru. Acesta raționalizează ordinele de eliminare din diferite state membre și acționează ca punct unic de contact în contextul semnalărilor și al ordinelor de eliminare și este menit să prevină duplicarea ordinelor de eliminare, mai exact același ordin trimis de două state membre. De asemenea, este util ca HSP să primească în mod centralizat ordinele de eliminare și semnalările primite de-a lungul timpului, pentru a sprijini obligațiile de

raportare în materie de transparență. Prin intermediul PERCI, HSP pot, de asemenea, să solicite control și revizuire pentru a contesta un ordin de eliminare.

3. Ce este autoritatea competentă a unui stat membru al UE și cum îi pot contacta?

Contactarea autorității competente cu privire la reprezentantul legal este necesară în anumite circumstanțe. O astfel de "circumstanță" apare atunci când se contestă un ordin de eliminare. Un alt caz ar fi, de exemplu, dacă HSP ia cunoștință de conținutul cu caracter terorist care implică o amenințare iminentă la adresa vieții sau un act terorist fără un ordin de eliminare. În acest caz, HSP are obligația de a-l șterge imediat și de a informa autoritatea competentă a statului membru al UE afectat de acesta (Regulamentul TCO, Art. 14.5).

Majoritatea autorităților competente din statele membre ale UE au înființat deja puncte de contact. O listă actualizată care include datele de contact este disponibilă pe site-ul Comisiei Europene aici.

Capitolul 5

Stabilirea unui mecanism de notificare și contestație pentru utilizatori în privința conținutului eliminat

Sinteză: Conținutul și principalele puncte ale prezentului capitol

- **Utilizatorii afectați pot apela la proceduri de contestație pentru a contesta ordinele de eliminare** (și, dacă este necesar și dincolo de domeniul de aplicare al TCO, alte măsuri proactive de moderare).
- Procedurile de prezentare a contestațiilor sunt importante ca **mecanism de control și feedback**, din perspectiva utilizatorilor, a companiilor și a legii.
- HSP "stabilesc un mecanism eficace și accesibil" de prezentare a contestațiilor (Regulamentul TCO, Art. 10.1).
- Conform Regulamentului TCO, mecanismele de prezentare a contestațiilor trebuie să îndeplinească anumite cerințe tehnice și de conținut.
- O procedură de contestație poate avea două rezultate diferite, și anume (1) contestația este admisă deoarece se constată că un anumit conținut a fost blocat în mod eronat sau (2) contestația este respinsă deoarece se constată că un anumit conținut a fost blocat în mod justificat.
- În funcție de rezultatul procesului de prezentare a contestațiilor, conținutul poate face obiectul altor măsuri.
- Acest capitol oferă, de asemenea, îndrumări privind conceperea și punerea în aplicare a unei proceduri de depunere a contestațiilor specifice HSP.

Procedurile de contestație permit utilizatorilor să conteste eliminarea conținutului prin comunicarea cu platforma și **reprezintă primul pas către contestarea (legală) a unui ordin de eliminare**. Puteți citi mai multe despre procesul de soluționare a litigiilor în ▶ capitolul 2.

1. De ce este necesar să se instituie un mecanism transparent de prezentare a contestațiilor?

Este important să se instituie un mecanism de prezentare a contestațiilor, în special din (a) perspectiva juridică, (b) perspectiva utilizatorului și (c) perspectiva societății.

a) Perspectiva juridică

Procedurile de depunere a contestațiilor sunt conforme cu **reglementările stabilite prin lege** care **fac necesare astfel de mecanisme**. De exemplu, Regulamentul TCO (Regulamentul TCO, Art. 10)

prevede că HSP trebuie să stabilească un mecanism eficace și accesibil de prezentare a contestațiilor pentru a oferi utilizatorilor posibilitatea de a contesta eliminarea sau blocarea conținutului după aplicarea unei anumite măsuri. În temeiul Regulamentului TCO, furnizorul de conținut trebuie să fie informat cu privire la rezultatul contestației în termen de două săptămâni.

b) Perspectiva utilizatorului

Un mecanism de depunere a contestațiilor nu este doar o cerință legală. Un mecanism clar și accesibil contribuie la consolidarea încrederii utilizatorilor, atât în rândul celor al căror conținut a fost moderat, cât și, în special, în rândul utilizatorilor neafectați care utilizează platforma conform destinației. În acest fel, HSP demonstrează că utilizatorii se pot baza pe procese de moderare, că astfel de procese se bazează pe o **responsabilitate față de utilizatori** și că platformele respectă **drepturile fundamentale**, cum ar fi libertatea de exprimare și de informare.

c) Perspectiva companiei

Mecanismele de depunere a contestațiilor pot fi o formă utilă de **automonitorizare**, prin care măsurile și standardele de moderare pot fi evaluate din punctul de vedere al eficacității, corectitudinii și consecvenței. Astfel de mecanisme pot oferi o asigurare că platforma dvs. este utilizată în conformitate cu scopul său, ceea ce contribuie și mai mult la protejarea **reputației** serviciilor online și la protejarea în continuare a dreptului la libertatea de exprimare online.

2. Care sunt cerințele unui mecanism de prezentare a contestațiilor?

În conformitate cu Regulamentul TCO (33 & Art. 10), mecanismele de prezentare a contestațiilor ar trebui:

- Să fie ușor de utilizat,
- Să fie eficace și (ușor) accesibile,
- Să asigure un cadru sigur în care contestațiile sunt tratate prompt și transparent, astfel încât reclamantul să fie informat cu privire la rezultatul analizei în termen de două săptămâni.

Procedurile de contestații trebuie instituite în sensul Regulamentului TCO pentru a restabili conținutul eliminat în mod eronat. Cu toate acestea, pot fi audiate contestații și împotriva măsurilor luate pentru a asigura respectarea ToS ale platformei dincolo de domeniul de aplicare al Regulamentului TCO.

3. Cum trebuie tratate contestațiile și care sunt posibilele rezultatele?

Odată ce mecanismul de contestații este implementat și utilizatorii au depus o contestație, HSP va **examina contestația** și **va comunica decizia reclamantului în termen de cel mult două săptămâni pentru cazurile care implică Regulamentul TCO**.

Există două rezultate posibile ale procesului de examinare, enumerate în tabelul următor.

Rezultatul A **Contestația utilizatorului împotriva eliminării conținutului a fost admisă**	*Rezultatul B* **Contestația utilizatorului împotriva eliminării conținutului a fost respinsă**
Rezultat: Contestația împotriva eliminării conținutului este **justificată** și, prin urmare, este admisă.	**Rezultat:** Contestația împotriva eliminării conținutului este justificată și, prin urmare, este respinsă.
Semnificație: Conținutul a fost eliminat, șters sau moderat în mod necorespunzător.	**Semnificație:** Conținutul a fost eliminat, șters sau moderat în mod corespunzător.
Procedura subsecventă: HSP (1) informează reclamantul cu privire la rezultatul analizei și (2) restabilește conținutul.	**Procedura subsecventă:** HSP (1) informează reclamantul cu privire la rezultatul analizei și (2) oferă utilizatorului motivele deciziei.

4. Îndrumări și sfaturi practice: Ce elemente sunt utile în stabilirea unui mecanism de prezentare a contestațiilor?

Faptul că un mecanism de contestații trebuie să aibă anumite caracteristici este o cerință nu doar a legii, ci și a **ușurinței de utilizare**.

Ca și alte aspecte-cheie, cum ar fi ToS (►capitolul 1), instituirea unui proces de identificare a conținutului interzis (►capitolul 2), sau alegerea mecanismelor de moderare specifice platformelor (►capitolul 3), metoda de depunere a unei contestații poate varia foarte mult de la o platformă la alta. În timp ce unele platforme pot oferi posibilitatea de le depune prin e-mail, altele aleg să implementeze aplicații standardizate bazate pe formulare online. **Mecanismele de contestații ar trebui să fie adaptate** scopului, structurii și organizării platformei. Orientări privind stabilirea unui mecanism de contestații pot fi găsite în următoarea **listă de verificare**.

Oferiți informații clare privind eliminarea conținutului

Informați persoana al cărei conținut a fost eliminat. Luați în considerare acest lucru chiar dacă conținutul a fost moderat în alt mod. În acest context, informați, de asemenea, persoana cu privire la motivul pentru care a fost eliminat conținutul (a se vedea mai jos următorul punct privind educația și cunoștințele pedagogice) și modul în care poate fi introdusă o cale de atac împotriva deciziei de eliminare (a se vedea următorul punct privind explicarea procesului de depunere a contestațiilor).

Explicați procedura de contestație

Explicați-le utilizatorilor dvs. procedura de contestație. Acest lucru ar trebui realizat odată cu notificarea faptului că un conținut a fost eliminat sau moderat în alt mod. Informații despre procedura de contestație pot fi, de asemenea, incluse în ToS. Explicațiile privind procedura de contestație ar trebui să includă: (1) modul de depunere a contestațiilor, (2) modul în care se realizează procesul de analiză și (3) modul în care utilizatorii sunt notificați cu privire la rezultatul analizei.

Furnizați informații pedagogice de bază

Oferiți utilizatorilor oportunități educaționale pentru a explica de ce a fost eliminat conținutul și ce termeni încalcă. De exemplu, conținutul poate fi eliminat dacă încalcă ToS sau dacă a fost primită o solicitare de eliminare de la o autoritate competentă în baza Regulamentului TCO. În primul caz, aceste informații pot fi specificate adăugând care dintre interdicțiile specifice platformei din ToS a fost încălcată (de exemplu, împotriva discursului de incitare la ură, a incitării la violență, a conținutului sexual și a hărțuirii). În cel din urmă caz, se recomandă adăugarea la procedura de contestații a unei scurte prezentări informative asupra Regulamentului TCO, astfel încât cadrul legal să fie clar pentru utilizator.

Oferiți actualizări periodice cu privire la evoluția contestației

Utilizatorii ar trebui să primească actualizări periodice cu privire la evoluția contestației, pentru a arăta că procesul este în desfășurare. Astfel de actualizări ar trebui să constea cel puțin într-o notificare atunci când analiza a fost finalizată și în comunicarea în timp util a rezultatului. O comunicare mai detaliată ar putea informa utilizatorul asupra primirii contestației și a faptului că este examinată de angajații HSP. Cele mai bune practici ar oferi utilizatorului un calendar pentru soluționarea contestației sale. Aceste actualizări pot fi date prin e-mail sau afișate într-un portal online.

Documentați procesul de contestație (individuală)

Este important să documentați procesul de contestație. Acest document va servi drept referință pentru orice întrebări sau dispute ulterioare.

Capitolul 6

Sprijin practic și consiliere privind rapoartele de transparență

Sinteză: Conținutul și principalele puncte ale acestui capitol

- Rapoartele de transparență permit HSP să **comunice public** modul în care **valorile** lor sunt respectate pe platformă, precum și **acțiunile întreprinse împotriva conținutului și comportamentului interzis și ilegal**.
- Rapoartele de transparență reprezintă un **instrument important pentru asumarea responsabilității publice și pentru demonstrarea credibilității**.
- Diferite **legi**, inclusiv Regulamentul TCO, impun în mod explicit **raportarea în mod transparent**.
- În pregătirea rapoartelor de transparență, se recomandă să se procedeze sistematic înainte, în timpul și după proces, urmând pași specifici.
- **Regulamentul TCO impune HSP care sunt expuși sau au luat măsuri împotriva conținutului cu caracter terorist să publice un raport anual de transparență cu privire la activitățile lor legate de conținutul cu caracter terorist**. Raportul trebuie să includă anumite cifre și informații esențiale, cum ar fi numărul de elemente de conținut eliminate. Raportul trebuie publicat până cel târziu la data de 1 martie a anului următor (Regulamentul TCO, Art. 7.2).

1. Ce sunt rapoartele de transparență?

Rapoartele de transparență reprezintă un **instrument important pentru HSP pentru a-și dovedi responsabilitatea și credibilitatea și pentru a publica informații relevante din punct de vedere social**. Rapoartele de transparență **conțin date importante cu privire la cererile pe care HSP le primesc de la actori statali din întreaga lume și cu privire la modul în care sunt tratate aceste cereri**, ceea ce este menit să asigure transparența colaborării și cooperării cu autoritățile și cu alte organisme de stat[7].

Rapoartele de transparență oferă, de asemenea, o imagine de ansamblu asupra măsurilor luate de HSP pentru a pune în aplicare reglementările (de exemplu, prin eliminarea conținutului sau alte măsuri de moderare). Aceasta include aplicarea (1) politicilor specifice platformei (de obicei, ToS; ▶capitolul 1), (2) drepturilor, cum ar fi legea drepturilor de autor sau a mărcilor comerciale și (3) legislației și reglementărilor (locale) care au ca rezultat eliminarea conținutului[8]. În UE, legislația

[7] Urman, A., & Makhortykh, M. (2023). How transparent are transparency reports? Comparative analysis of transparency reporting across online platforms. *Telecommunications Policy, 47*(3), 102477. https://doi.org/10.1016/j.telpol.2022.102477

[8] Trust & Safety Professional Association (2023). What Is A Transparency Report? *TSPA*. https://www.tspa.org/curriculum/ts-fundamentals/transparency-report/what-is-a-transparency-report/

locală include Regulamentul TCO și Actul privind serviciile digitale. Legislația specifică fiecărei țări poate fi, de asemenea, relevantă, cum ar fi Legea privind rețelele din Germania.

Rapoartele de transparență sunt, de regulă, publicate în mod regulat. Regulamentul TCO prevede că acest lucru ar trebui făcut **(cel puțin) o dată pe an** dacă un HSP a luat măsuri împotriva conținutului cu caracter terorist (Regulamentul TCO, 30).

De la HSP la HSP, rapoartele de transparență și valorile raportate în acestea pot varia foarte mult[9]. **Pentru rapoartele privind transparența în conformitate cu Regulamentul TCO, există cerințe specifice cu privire la informațiile care trebuie incluse.** Puteți citi mai multe despre acest lucru în ►sectiunea 4 din acest capitol.

2. De ce sunt necesare rapoartele de transparență?

Rapoartele de transparență **permit utilizatorilor și terților să evalueze măsura în care HSP continuă să respecte propriile principii, cerințele legale, precum și protecția datelor și a vieții private**[9]. Să analizăm diferitele perspective asupra importanței rapoartelor de transparență.

a) Perspectiva legală

Rapoartele de transparență sunt utile și adesea necesare pentru a respecta reglementările UE, cum ar fi Regulamentul TCO, Actul privind serviciile digitale și legile naționale, acolo unde este cazul. În cazul în care HSP au luat măsuri împotriva diseminării conținutului cu caracter terorist în cursul unui an calendaristic, fie în mod proactiv, fie în conformitate cu un ordin de eliminare, **trebuie publicat un raport de transparență până cel târziu la data de 1 martie a anului următor** (Regulamentul TCO, Art. 7.2). Aceasta înseamnă că este mai probabil ca rapoartele privind transparența să fie obligatorii decât să nu fie. Autoritățile competente au, de asemenea, obligația de a publica rapoarte anuale privind transparența (Regulamentul TCO, 31).

b) Perspectiva utilizatorilor și a părților interesate

Rapoartele de transparență ajută utilizatorii și alte părți interesate să evalueze măsura în care HSP își îndeplinesc **responsabilitatea față de societate**. În același timp, publicarea periodică a rapoartelor de transparență ajută HSP să **consolideze încrederea și să-și construiască o bună reputație publică**, prin demonstrarea conformității, angajamentului și fiabilității.

c) Perspectiva companiei

Similar mecanismelor de depunere a contestațiilor, rapoartele de transparență pot fi, de asemenea, un **instrument de automonitorizare** care identifică domeniile în care procesele pot fi optimizate. Având în vedere că HSP mai mici sunt mai populari în rândul actorilor teroriști (pentru detalii, a se vedea acest raport Tech Against Terrorism), este de datoria acestora să ia măsuri împotriva

[9] Woolery, L., Budish, R., & Bankston, K. (2016). The transparency reporting toolkit. *New America and The Berkman Center for Internet & Society at Harvard University*.

diseminării conținutului terorist. Rapoartele privind transparența reprezintă o modalitate prin care HSP mai mici își pot demonstra angajamentul față de acest efort.

3. Procesul de pregătire a rapoartelor de transparență

Întocmirea și livrarea inițială a unui raport de transparență poate fi o sarcină descurajantă. Cu toate acestea, odată ce au fost stabilite **un proces și o rutină pentru pregătirea rapoartelor anuale de transparență**, de obicei nu este necesară refacerea completă a structurii raportului, iar actualizările sunt de obicei suficiente. Acest ghid va sublinia modul de întocmire a unui raport de transparență din perspectiva HSP, precum și ce trebuie luat în considerare în avans, în timpul și după procesul de creare.

a) **Înainte** de întocmirea raportului de transparență

Scanați peisajul juridic

Obțineți o imagine generală a reglementărilor legale care se aplică HSP. De exemplu, Regulamentul TCO și Actul privind serviciile digitale sunt relevante în întreaga UE. Pot exista reglementări specifice fiecărei țări, precum și cerințe pentru alte subiecte relevante pentru HSP, în afara conținutului cu caracter terorist. Dacă ați angajat persoane de contact sau avocați, este adesea recomandabil să discutați cu aceștia pentru a determina ce reglementări (în afară de Regulamentul TCO cu care aveți de-a face aici) se aplică HSP.

Stabiliți obiectivele raportului de transparență

Gândiți-vă care sunt obiectivele raportului dvs. de transparență. Întrebările cheie care vă pot ajuta în acest sens sunt: Doriți pur și simplu "doar" să vă îndepliniți obligațiile legale sau doriți să abordați și alte subiecte și angajamentul dvs. față de acestea? Cui doriți să vă adresați, adică ce grup țintă are HSP (de exemplu, actori politici, utilizatori, finanțatori)? Cât de des doriți să publicați rapoarte privind transparența și care este cel mai bun moment pentru exercițiul financiar individual?

Determinați ce date pot și ar trebui incluse

Stabiliți ce date veți include în raportul de transparență. Pe de o parte, *capacitatea* este relevantă în acest scop, adică ce date sunt disponibile sau care dintre acestea este fezabil să le puteți colecta în viitor? Pe de altă parte, *ceea ce este obligatoriu* este decisiv, adică ce cerințe legale trebuie să îndepliniți și ce date sunt necesare pentru aceasta? Puteți afla ce informații și date trebuie să includeți în conformitate cu Regulamentul TCO în ► subcapitolul următor.

b) **În timpul** elaborării raportului de transparență

Folosiți un limbaj clar și concis

În timp ce creați raportul de transparență, asigurați-vă că este utilizat un limbaj clar și concis. Acest lucru ajută la implicarea și înțelegerea materialelor complexe. Limba ar trebui adaptată în continuare la cititorii avuți în vedere.

Furnizați informații contextuale și explicații

Oferiți cititorilor informații contextuale și explicații. Astfel de explicații permit utilizatorilor să înțeleagă mai bine modul în care funcționează HSP și motivele pentru care (poate) oferă un raport de transparență mai detaliat decât "doar" un raport care îndeplinește doar cerințele minime de reglementare.

Încorporați feedback intern

Când programați un raport de transparență, încorporați runde de feedback în calendarul de întocmire. Feedback-ul regulat care oferă corecții conținutului și limbajului raportului poate fi valoros pentru toți actorii implicați, în special pentru cei însărcinați cu pregătirea raportului.

c) **După** întocmirea raportului de transparență

Publicați raportul de transparență

Gândiți-vă la limbile în care doriți să publicați raportul de transparență și realizați traduceri adecvate. De asemenea, determinați unde ar trebui să fie accesibil, adică secțiunea de pe site-ul dvs. În plus, puteți lua în considerare includerea raportului de transparență în diverse materiale de comunicare, pentru a atrage mai multă atenție asupra sa. Aceasta ar putea include încorporarea acestuia pe website-ul dvs., trimiterea acestuia în buletine informative prin e-mail sau partajarea acestuia pe social media.

Actualizați periodic datele și, în cele din urmă, raportul de transparență

Când primul raport de transparență a fost finalizat, ați construit deja fundația care facilita elaborarea următorului, având în vedere că ați creat deja un șablon. Procesul de colectare a datelor relevante pe tot parcursul anului și organizarea acestora pentru a le face ușor accesibile vor face ca elaborarea următorului raport de transparență să fie mai ușoară și să necesite mai puțin timp și resurse.

Lăsați loc de îmbunătățire

Fiți deschiși la schimbări și ajustări. Dacă primiți feedback extern, luați în considerare includerea acestuia în următorul raport de transparență, dacă este cazul. Cu toate acestea,

feedback-ul relevant nu poate fi doar extern. După publicare, HSP poate, de asemenea, să arunce o privire critică asupra raportului anterior de transparență, a comunicării din jurul acestuia și a reacțiilor la acesta, luând în considerare declarațiile de presă sau alte surse relevante. Atât feedback-ul intern, cât și cel extern pot identifica domeniile în care pot fi aduse îmbunătățiri.

4. Ce informații și indicatori trebuie incluși în raportul de transparență?

Regulamentul TCO explică la Articolul 7.3 **cerințele minime pentru rapoartele de transparență**, și anume exact ceea ce trebuie inclus pentru a respecta această legislație a UE. Prezentăm aceste cerințe într-o listă de verificare de mai jos:

1) **Informații** privind măsurile luate de HSP:

- ☑ identificarea conținutului cu caracter terorist;
- ☑ eliminarea sau dezactivarea conținutului cu caracter terorist;
- ☑ prevenirea reapariției și reîncărcării materialelor online blocate anterior (acest lucru este deosebit de relevant atunci când se utilizează proceduri automatizate).

2) **Indicatori** și, dacă este cazul, informații suplimentare privind numărul de:

- ☑ elemente eliminate care includ conținut cu caracter terorist (pe baza ordinelor de eliminare sau a altor măsuri);
- ☑ ordinele de eliminare în urma cărora nu s-a acționat și informații suplimentare cu privire la motivele pentru care s-a procedat astfel;
- ☑ contestațiile prelucrate de HSP prin intermediul mecanismului de tratare a contestațiilor, precum și informații suplimentare privind rezultatul contestațiilor;
- ☑ cazurile în care HSP a restabilit conținutul în urma contestației furnizorului de conținut;
- ☑ procedurile judiciare inițiate de HSP și informații suplimentare cu privire la rezultatul acestora;
- ☑ cazurile în care HSP a trebuit să restabilească conținutul în urma procedurilor judiciare.

C. Mulțumim pentru ajutorul acordat în combaterea amenințării teroriste!

Felicitări!

Ați ajuns până aici, ceea ce înseamnă că ați dobândit cunoștințe esențiale cu privire la cerințele Regulamentului TCO și la măsurile suplimentare de combatere a conținutului online dăunător și a altor conținuturi dăunătoare. În acest fel, veți contribui la un internet mai sigur.

Suntem conștienți de faptul că punerea în aplicare a acestor măsuri necesită o atenție și resurse considerabile. Faptul că v-ați angajat cu acest ghid este o mișcare excelentă și dacă acest ghid vă ajută să vă gândiți la o strategie de implementare a măsurilor împotriva conținutului terorist care este potrivită pentru HSP-ul dvs., atunci am realizat deja multe! Suntem încrezători că acest ghid, alături de celelalte resurse educaționale ale noastre, va fi un instrument neprețuit pentru a permite HSP să își îndeplinească pe deplin obligațiile în ceea ce privește combaterea unei amenințări teroriste online.

Vă mulțumim foarte mult pentru angajamentul dumneavoastră de a contracara amenințarea teroristă online și pentru că vă apărați HSP și utilizatorii.

PS: Proiectul Tech Against Terrorism Europe include, de asemenea, un **curs online gratuit, premiat**, pentru a explora în continuare conformitatea cu Regulamentul TCO al UE. În acest curs, puteți aprofunda subiectul și puteți găsi detalii despre regulament, exemple despre modul în care alte platforme pun în aplicare măsuri individuale și mai multe informații generale privind comportamentul terorist online. **După finalizarea cu succes a cursului, veți primi un certificat oficial semnat de universități de renume** (LMU München, Universitatea din Gent). În plus, TATE oferă un program de consolidare a capacităților în care HSP pot primi sprijin practic pentru cerințele legate de Regulamentul TCO.

D. Glosar

Termen	Explicație
Autorități competente	Autoritățile unui stat membru al UE care sunt responsabile pentru punerea în aplicare a Regulamentului TCO. O prezentare generală a autorităților competente respective ale statelor membre poate fi găsită aici.
Furnizor de conținut	Persoana care furnizează conținut pe platforma respectivă, de exemplu, publică o postare.
HSP	Furnizori de servicii de găzduire; Regulamentul TCO se aplică HSP. Mai multe detalii despre HSP care intră în domeniul de aplicare al Regulamentului TCO pot fi găsite ►aici.
PERCI	PERCI este un instrument coordonat de Europol menit să îmbunătățească și să faciliteze comunicarea dintre HSP și autoritățile competente.
Ordin de eliminare	O cerere pe care un HSP o primește de la o autoritate competentă. Aceasta informează HSP că pe platformă a circulat conținut cu caracter terorist și obligă HSP să îl elimine rapid, în termen de o oră de la primire.
Regulamentul TCO (de asemenea: LEX 2021/784 & Regulamentul privind combaterea diseminării conținutului online cu caracter terorist)	Regulamentul UE împotriva diseminării conținutului online cu caracter terorist, care a intrat în vigoare în 2022. Regulamentul se aplică furnizorilor de servicii de găzduire (HSP) care își oferă serviciile în UE.
Conținutul online cu caracter terorist (TCO)	Conținut care include elemente teroriste sau care promovează scopuri teroriste. O definiție detaliată a acestuia poate fi găsită în ►introducere. Conținutul online cu caracter terorist este strâns legat de ►infracțiunile teroriste, care pot avea un caracter foarte divers.
ToS	Clauze și condiții; Reguli obligatorii stabilite de și pentru platforma individuală care (1) definesc domeniul de aplicare și responsabilitatea HSP față de utilizatori și (2) practici de utilizare adecvate și permise, dar și interzise. Utilizatorii trebuie să le respecte dacă doresc să continue să utilizeze serviciul oferit de HSP. Există numeroase sinonime pentru ToS, de ex. Termeni de utilizare, Termeni și condiții sau Standarde comunitare