



БЪЛГАРСКИ  
ХЕЛЗИНСКИ  
КОМИТЕТ

## Препоръки

### относно гарантирането на свободата на изразяване във виртуална среда при предприемането на мерки спрямо речта на омразата и дезинформацията

#### Въведение:

Свободата на изразяване е част от фундамента на демократичното общество. Тя се отнася не само до информация и идеи, които намират благоприятен прием или се смятат за безвредни, но и до такива, които шокират, смущават или разстройват държавата или отделни сегменти на обществото. Такива са изискванията на плурализма и широкоумието, без които няма демократично общество.

Същевременно, надхвърлянето на допустимите граници на свободата на изразяване, под формата на речта на омразата, дезинформация и манипулиране на общественото мнение са заплаха за демократичните ценности и мира. Наблюдаваме тревожни тенденции на реч на омразата, дезинформация и манипулиране на общественото мнение във виртуална среда, както и неприемливи мерки срещу тях от страна на правителства и компании, нарушаващи международните норми за правата на човека. В отговор на тези тенденции се налага преразглеждане и актуализация на политиките, гарантиращи свободата на изразяване във виртуална среда и ограничаване на злоупотребите с нея. Този процес следва да се основава на формулираните по-долу препоръки, които обобщават законите, политиките и добрите практики, възприети или в процес на възприемане в Европейския съюз.

#### Препоръки:

1. Човешките права, които хората имат извън виртуална среда, трябва да бъдат защитени и във виртуална среда. Това се отнася особено за гражданските свободи, включително правото на неприкосновеност на личния живот, правото на изразяване, религия, сдружаване и др.
2. Свободата на изразяване във виртуална среда е същата както и извън такава среда.
3. Свободата на изразяване във виртуална среда, подобно на тази във всяка друга среда, подлежи на формалности, условия, ограничения или санкции, които са предвидени от закона и са необходими в едно демократично общество в интерес на националната сигурност, териториалната цялост или обществената безопасност, за предотвратяване на безредици или престъпления, за защита на здравето или морала, за защита на репутацията или правата на други лица, за предотвратяване на разкриването на поверителна информация или за поддържане на авторитета и безпристрастността на съдебната власт.
4. Държавата има задължението да уважава, защитава и способства за реализирането на правата на човека, което предполага и да приема политики и регулации, осигуряващи свободата на изразяване във виртуална среда и нейното легитимно ограничаване. Всички организации, включително бизнес компании, и всички граждани имат задължението да уважават правата на човека във виртуална среда.
5. При изработването, приемането и развитието на политики и нормативни механизми държавата трябва да осигури достъп до проектите за такива документи, прозрачност и ефективно участие на гражданите, чрез открит и приобщаващ процес на консултации и/или обществено обсъждане.
6. Докато речта на омразата сама по себе си не е дефинирана и не е забранена в международното право по правата на човека, тя е незаконна при определени условия – когато **подбужда** към дискриминация, враждебност или насилие на основата на защитен

признак. Държавата трябва да създаде правни средства за търсене на отговорност за такова подбуждане. Реч на омразата, която не достига прага на подбуждане, макар да не е забранена сама по себе си, също може да бъде вредна и следва да получава отпор чрез подходящи образователни мерки.

7. Държавата следва да се противопоставя на разпространението на дезинформация, когато тя нарушава правата на човека, чрез подходящи политики, включително изграждане на капацитет за превенция и устойчивост срещу формите на дезинформация и повишаване на медийната грамотност на всички граждани.
8. При употреба на понятието „дезинформация“, държавата трябва да се стреми то да бъде пределно ясно и да не допуска широки тълкувания. Общоприетото определение на дезинформацията като доказуемо невярна информация, която се разпространява умишлено, с цел да причини сериозна социална или индивидуална вреда, трябва да се разглежда като граница, вътре в която съдържанието може да се конкретизира, но отвъд която не трябва да се преминава. Не представляват дезинформация неумишлени грешки, сатири, пародии. Не са дезинформация сами по себе си ясно идентифицирани политически пристрастни, или пропагандни новини и коментари. Не може да се определи като дезинформация дадено съдържание само поради това, че е невярно, обидно или вредно; или само защото може да провокира общественото мнение; или само защото може да накърни общественото спокойствие или националната икономика, да засегне репутацията на публични институции, публични фигури и други участници в обществения живот; или само защото представлява слухове, митове или „конспиративни теории“.
9. Сами по себе си враждебното и неистинното съдържание, включително дезинформация, не са незаконни. Те са включени в обхвата на защитеното изразяване и могат да се ограничават и санкционират само ако попадат в сферата на легитимните ограничения, посочени в закона. Обща забрана за изразяване на неистини, погрешни мнения или неправилни тълкувания на минали събития не е допустима.
10. Мерките срещу разпространението на дезинформация трябва да са в съответствие с международните норми за правата на човека и принципите на законност, необходимост и пропорционалност и да подчертават значението на свободните, независими, плуралистични и разнообразни медии и на предоставянето и насърчаването на достъп до независима, основана на факти и доказателства информация.
11. Държавата трябва да възпира държавните служители да разпространяват реч на омраза или дезинформация, и за тази цел трябва да създаде критерии в професионални кодекси на поведение.
12. Всяко ограничение, което се счита за оправдано поради това, че подбужда към дискриминация, враждебност или насилие на основата на защитен признак, или което е дезинформация, трябва да бъде в съответствие и с общите принципи на ограничаване на свободата на изразяване, посочени по-горе в Принцип 3.
13. При ограничаването на свободата на изразяване поради подбуждане към дискриминация, враждебност или насилие въз основата на защитен признак, трябва да се прави конкретна оценка за необходимостта и пропорционалността на ограничителната мярка чрез анализ на следните фактори: контекст на изявлението, наличието на обществен дебат по темата, статус на говорещия, статус на този, към когото е насочено изразяването, намерение да се навреди на уязвима група, съдържание и форма на изразяването, степен на разпространението му, и вероятност от нанасяне на вреда.
14. При приемането и прилагането на политики и регулации, осигуряващи свободата на изразяване във виртуална среда и нейното легитимно ограничаване, държавата трябва да осигури защитата на личните данни и неприкосновеността на личния живот в съответствие с действащото законодателство.

15. Държавата трябва да гарантира спазването на забраната за целева реклама във виртуална среда, основана на профилиране посредством чувствителни данни по смисъла на член 9 от GDPR: данни за расов или етнически произход; политически възгледи, религиозни или философски убеждения; членство в синдикални организации; генетични данни; биометрични данни за целите на еднозначното идентифициране на физическо лице; данни за здравословното състояние, сексуалния живот или сексуалната ориентация на физическо лице.
16. Държавата трябва да предприеме мерки за осигуряване на прозрачност на политическата реклама във виртуалното пространство, включително да изисква посочване на факта, че съдържанието е политическа реклама, както и нейното авторство.
17. Държавата трябва да установи механизми за независим надзор на технологичните компании, насочен към защита на човешките права и правна защита за потребителите на тези компании, при стриктно спазване на свободата на изразяване. При въвеждането на подобни механизми държавата трябва да уважава легитимните изисквания за търговска конфиденциалност, и да диференцира мерките в зависимост от размера и структурата на компаниите.
18. Всички публични и частни създатели на изкуствен интелект имат задължението да осигурят всеки изкуствен интелект да се идентифицира като такъв по ясен начин преди всяко взаимодействие с потребител, което съдържа реклама на продукт или може да влияе върху индивидуалното поведение на потребителя в сферата на политическото му участие.
19. Технологичните компании, които предоставят цифрови услуги, трябва:
  - a. Да предприемат мерки за надлежна проверка, за да идентифицират **незаконно съдържание** и дейности на своите платформи, и да противодействат на речта на омразата, незаконната дезинформация и манипулативното поведение във виртуална среда.
  - b. В случай на незаконно съдържание, да прилагат политики и инструменти за модерирание на съдържанието, включително мерки като премахване, етикетирание, спиране на акаунти, демонетизация и деамплификация на съдържание, санкциониране на кликбеята, инициативи за проверка на факти в партньорство с доверени сигнализатори, и др. Тези мерки следва да отговарят на изискванията за необходимост и пропорционалност.
  - c. Да организират механизъм за уведомяване, позволяващ на всяко физическо или юридическо лице да съобщава за определена информация, която лицето счита за незаконно съдържание.
  - d. Да осигуряват приоритетно разглеждане на уведомленията, подадени от субекти, наричани "доверени сигнализатори" (trusted flaggers), които работят в определени области, в които имат експертен опит.
  - e. Да приемат и прилагат политики за противодействие на **незаконно поведение** във виртуална среда, например чрез фалшиви акаунти, усилване на информацията чрез ботове, олицетворяване, тролови ферми, злонамерени дълбоки фалшификати – deepfake, и др.
  - f. Да осигуряват прозрачност на своите политики и процедури и да предоставят подробна информация за това как се предприемат ограничителните мерки в различни географски области и на различни езици.
  - g. Да предоставят на специално определен регулаторен орган периодични доклади за предприетите действия за справяне с незаконно съдържание и поведение; в тях да се включва, например, информация за броя на заповедите, които са им били изпратени от държавни органи и организации във връзка с модерирание на съдържание или искана от тях информация за потребители; обобщена информация за човешките ресурси, ангажирани с модерирание на съдържание; броя на акаунтите и единиците

съдържание, които са били свалени от компанията; точността и степента на грешка на техните автоматизирани системи за модерирание на съдържание.

- h. Да осигуряват независим мониторинг и измерване на въздействието на ограничителните мерки.
- i. Да осигуряват достъп на изследователи до съответните данни и информация, за да могат те да разберат динамиката, свързана с речта на омразата и дезинформацията, и да спомагат създаването на основани на доказателства политики и регулаторни мерки.
- j. Да поддържат независими механизми за разглеждане на жалби, включително процедури за обжалване или преразглеждане на премахването на съдържание, и предоставяне на ефективни средства за защита на потребителите.
- k. Постоянно да въвеждат и усъвършенстват функционалности, които дават на потребителите по-голям контрол във виртуална среда, включително механизми за разрешаване на спорове, прозрачност по отношение на решенията за модерирание на съдържанието и възможности за оспорване на решенията на платформата от страна на потребителите.
- l. Да провеждат надлежна проверка по отношение на ефектите от техните продукти, политики и операции, в съответствие с техните отговорности съгласно Ръководните принципи на ООН за бизнеса и правата на човека.
- m. Компании, които са много големи онлайн платформи (VLOP) или много големи онлайн търсачки (VLOSE) по смисъла на Регламент (ЕС) 2022/2065 на Европейския парламент и на Съвета от 19 октомври 2022 година (Digital Services Act), имат допълнителното задължение да оценяват и предприемат мерки за преодоляване на системни рискове, произтичащи от проектирането, функционирането и използването на техните услуги, както и от потенциалната злоупотреба с услуги от страна на получателите на услуги; при оценката на системните рискове те следва да се съсредоточат върху системите или други елементи, които могат да допринесат за риска, включително всички алгоритмични системи, които могат да бъдат от значение, по-специално техните системи за препоръки и реклама.
- n. Системни рискове включват, между други, а) рискът, свързан с действително или предвидимо отрицателно въздействие върху демократичните процеси, гражданския дискурс и изборните процеси, както и върху обществената сигурност; б) рискът, свързан с действителен или предвидим отрицателен ефект върху защитата на общественото здраве, непълнолетните, и сериозни отрицателни последици за физическото и психическото благополучие на лице, или насилие, основано на пола. Такива рискове могат също така да произтичат от координирани кампании за дезинформация, свързани с общественото здраве, или от дизайн на онлайн интерфейс, който може да стимулира поведенчески пристрастявания на получателите на услугата.