БЪЛГАРСКИ НАЦИОНАЛЕН ДОКЛАД

ЗАЕДНО СРЕЩУ АНТИРОМСКИТЕ НАГЛАСИ ОНЛАЙН

София, 2026

БЪЛГАРСКИ НАЦИОНАЛЕН ДОКЛАД

ЗАЕДНО СРЕЩУ АНТИРОМСКИТЕ НАГЛАСИ ОНЛАЙН

TOGETHER AGAINST ANTIGYPSYISM ONLINE

THE PROJECTS IS FUNDED BY EUROPEAN COMMISION TROUGH THE CITIZENS EQUALITY RIGHTS AND VALUES

Co-funded by the European Union

Проект: Together Against Antigypsyism Online (TAAO)
Програма: Citizens, Equality, Rights and Values Programme (CERV-2023-EQUAL)
Номер на проекта: ----------
Продължителност: март 2024 – февруари 2026
Водеща организация: Amaro Drom e.V. (Germany)
Национален партньор: Integro Association, Bulgaria
Държави, включени в проекта: Белгия, България, Чехия, Унгария, Словакия, Румъния

Автори: К. Макавеев
В. Ибрям

София, 2026

Съдържание

**ПРЕДГОВОР И КОНТЕКСТ –
РОЛЯТА НА АСОЦИАЦИЯ ИНТЕГРО** 5

ОБОБЩЕНИЕ НА ПРОЕКТА ТААО – БЪЛГАРИЯ 7

ВЪВЕДЕНИЕ 9
1.1 Предистория и контекст 9
1.2 Значение на изследването 10
1.3 Обучение и изграждане на капацитет 10
1.4 Законодателна и нормативна рамка 11

МЕТОДОЛОГИЯ 15
2.1 Цел и дизайн 15
2.2 Участници - Българският екип за мониторинг 16
2.3 Инструменти и процедури на мониторинга 18

РЕЗУЛТАТИ 21
3.1 Количествен компонент на изследването 21
3.1.1 Резюме 21
3.1.2 Количествени анализи 22
A. Обща информация – общ брой докладвани случаи
и социални мрежи: типове акаунти и профили 22
B. Основна тема и представяне 25
C. Тип и стил на съдържанието 30
D. Нетекстови елементи и визуални материали 37

E. Интензивност и ниво на омразата -
екология и степен на токсичност 41
F. Противодействия и емоционални реакции..... 47
3.2 Качествен компонент на изследването 50
3.2.1 Основни тематични модели 51
3.2.2 Анализ на избрани случаи 53
3.2.3 Емоционални ефекти, онлайн динамика
и ключови наблюдения от мониторинга
на антиромската омраза в България..... 59

ДИСКУСИЯ 61
4.1 Интерпретация на резултатите 61
4.2 Значение и препоръки за
заинтересованите страни 62
4.3 Европейски контекст 63
4.4 Ограничения 63
4.5 Ключови изводи 64

Литература 66

Приложение 1: Общ инструмент за мониторинг 68

ПРЕДГОВОР И КОНТЕКСТ – РОЛЯТА НА АСОЦИАЦИЯ ИНТЕГРО

Асоциация Интегро е една от най-утвърдените ромски неправителствени организации в България. Основана през 2002 г., тя обединява екип от ангажирани професионалисти, работещи за преодоляване на дълбоко вкоренените предизвикателства пред ромската общност – стереотипи, предразсъдъци, социално изключване, сегрегация, дискриминация и специфична форма на расизъм, известна като антиромски расизъм (antigypsyism), които продължават да съществуват както в България, така и в цяла Европа.

Социалната маргинализация на ромите се подхранва и задълбочава от негативните обществени нагласи, които се възпроизвеждат и усилват чрез медийните наративи и публичния дискурс. В началото на 2000-те години, когато печатните медии, радиото и телевизията бяха основните източници на информация, Асоциация Интегро систематично анализираше начина, по който ромските общности се представят в публичните медии, последователно отстоявайки принципа, че свободата на словото не може да служи като оправдание за нарушаване на човешките права и че защитата на достойнството на всеки човек е основополагащ принцип на демократичното общество.

Водена от убеждението, че устойчивата промяна започва с младите хора, през 2010 г. Асоциация Интегро реализира в България европейската кампания REACT на ERGO Network под мотото „Благодарим Ви, г-н кмете!". Тази инициатива постави началото на първото систематично изследване на представянето на ромите в печатните медии, проведено в рамките на 12 месеца от осем млади ромски специалисти и студенти. Резултатите, публикувани през 2011 г., документираха трайното присъствие на негативни и стереотипни

образи на ромските общности и поставиха основата за последващи застъпнически кампании срещу речта на омразата в медиите.

През следващите години Интегро разшири своя мониторинг, включвайки радиото и телевизията. В рамките на проекта Curbing Anti-Gypsyism from Local to European Level, финансиран от Open Society Foundation - Budapest, през 2015 г. беше публикуван докладът „Образът на ромите в шест електронни медии". Този опит потвърди, че медийните наративи могат да се променят в положителна посока, когато младите хора активно се ангажират със защитата на човешките права.

С развитието на дигитализацията речта на омразата все по-често се пренася в онлайн пространството, където се разпространява значително по-бързо и допълнително засилва дискриминационните нагласи. В отговор на тази тенденция, през 2020 г. Интегро се включи в проекта Peer Education to Counter Antigypsyist Online Hate Speech (PECAO), координиран от ERGO Network. Надграждайки постигнатото в рамките на PECAO, настоящият проект Together Against Anti-Gypsyism Online (TAAO) обединява млади хора от шест европейски държави в активно противодействие срещу онлайн речта на омразата.

В България в периода октомври 2024 г. - септември 2025 г. национален екип от млади роми наблюдава и анализира 528 случая на онлайн реч на омразата. Всички участници преминаха двуседмично онлайн обучение, водено от доц. д-р Илеана Ротару (Западен университет в Тимишоара), в рамките на което придобиха умения за идентифициране, класифициране и анализ на съдържание, съдържащо реч на омразата, в съответствие с етичните и методологическите стандарти на проекта.

Екипът за мониторинг на Интегро се състоеше от петима млади специалисти под ръководството на Камен Макавеев (координатор на Работен пакет 4) и Вергил Ибрям (ръководител на мониторинга по Работен пакет 2). Екипът провеждаше редовни работни срещи, осигуряваше методологическа подкрепа и обратна връзка на наблюдателите и обобщаваше резултатите в тримесечни аналитични доклади. Настоящият доклад представя количествен и качествен анализ на събраните данни, като откроява активната роля на младите наблюдатели в противодействието на онлайн речта на омразата и насърчаването на социалното приобщаване.

ОБОБЩЕНИЕ НА ПРОЕКТА ТААО – БЪЛГАРИЯ

Проектът Together Against Anti-Gypsyism Online (TAAO), финансиран по програмата Citizens, Equality, Rights and Values (CERV-2023-EQUAL), е насочен към една от най-устойчивите и недостатъчно изследвани форми на расизъм в Европа - антиромския расизъм (antigypsyism). Проектът има за цел да изследва как речта на омразата срещу ромите се разпространява в дигиталните платформи, как се проявява сред потребители и медийни източници и как онлайн платформите реагират на подобно съдържание.

Проектът ТААО се координира от Amaro Drom e.V. (Германия) и се реализира в шест държави членки на ЕС: България, Чехия, Германия, Унгария, Румъния и Словакия. Той е в съответствие с приоритетите на Европейския съюз за защита на основните права и ценности чрез противодействие на речта на омразата и престъпленията от омраза, със специален фокус върху борбата с антиромския расизъм в онлайн среда.

В рамките на проекта национални екипи от най-малко пет обучени наблюдатели във всяка държава систематично наблюдаваха, документираха и анализираха случаи на антиромска реч на омразата, прилагайки единна европейска методология. В България екипът идентифицира и класифицира 528 случая в периода октомври 2024 г. - септември 2025 г., обхващащи социални мрежи, новинарски портали и други дигитални източници. Събраните данни бяха структурирани по тематичен фокус, вид на дискурса, степен на агресивност и реакция на съответната платформа.

Българският принос включва четири обобщени доклада, съответстващи на четирите периода на мониторинг, с общо 528 попъл-

нени формуляра Monitoring Tool (MT) - основният изследователски инструмент на проекта. Структурата и начинът на прилагане на MT са подробно представени в Приложение 1.

Анализът подчертава устойчивостта на антиромските внушения, които възпроизвеждат негативни стереотипи, подкопават ромската идентичност и нормализират социалното изключване. Много онлайн публикации използват кодирани послания, ирония или хумор, за да прикрият дискриминационни внушения - механизъм, който позволява на речта на омразата да остане социално толерирана и до голяма степен незабелязана. Констатациите показват също значителни различия в реакциите на онлайн платформите - от незабавно премахване на съдържанието до пълна липса на действия. Чрез предоставянето на систематизирани национални данни, съчетани със сравнителни анализи, проектът ТААО допринася за по-задълбочено разбиране на динамиката на онлайн речта на омразата и на начините, по които антиромският расизъм се адаптира към развиващите се форми на дигитална комуникация.

Тези изводи имат практическо значение за създателите на политики, образователните специалисти и организациите на гражданското общество - както ромски, така и неромски - като подпомагат разработването на по-ефективни стратегии за превенция, докладване и противодействие на онлайн речта на омразата.

Настоящият национален доклад е изготвен от Камен Макавеев с подкрепата на българския екип за мониторинг и доц. д-р Илеана Ротару (Западен университет в Тимишоара). Специална благодарност се изказва към членовете на екипа - Ахмед Хасанов, Вергил Ибрям, Гамзе Максимова, Диян Данков и Мария Христова - за техния професионализъм и отдаденост. Цялостната координация на проекта и управлението на качеството бяха осигурени от Лилия Макавеева, изпълнителен директор на Асоциация Интегро.

ВЪВЕДЕНИЕ

1.1 Предистория и контекст

Настоящият български национален доклад е част от многонационалното изследване по проекта Together Against Antigypsyism Online (TAAO), реализирано паралелно в шест държави членки на ЕС. Докладът обхваща периода октомври 2024 г. - септември 2025 г. и е изготвен от Асоциация Интегро въз основа на систематичен мониторинг на онлайн съдържание, насочено срещу ромската общност в България.

Мониторингът беше организиран около четири отчетни периода, като данните бяха събирани непрекъснато и систематично. Националният екип провеждаше редовни вътрешни срещи за обсъждане на трудни случаи, изясняване на методологически въпроси и осигуряване на последователност в оценките. Едновременно с това ежемесечните международни координационни срещи с представители от всички партньорски организации позволиха уеднаквяване на подходите и споделяне на опит между националните екипи.

Целта на настоящия доклад е да представи по задълбочен и структуриран начин констатациите от мониторинга - не само като съвкупност от статистически данни, но и като аналитична рамка, която разкрива механизмите, формите и въздействието на онлайн антиромския дискурс в България. Изводите имат практическо значение за граждански организации, институции и платформи, ангажирани с противодействието на речта на омразата.

1.2 Значение на изследването

През последните години използването на интернет, социалните мрежи и дигиталните платформи нараства с изключително бързи темпове. Макар това развитие да създава значителни възможности, то същевременно излага хората - включително деца и младежи - на различни форми на онлайн реч на омразата, независимо дали като свидетели, участници или жертви. В този контекст ромската общност остава една от най-силно засегнатите групи, като проявите на дискриминация и стереотипизиране могат да доведат до сериозни социални и психологически последици.

Проектът ТААО надгражда усилията на предходния проект РЕСАО (Peer Education to Counter Antigypsyist Hate Speech Online) и е насочен към системното проследяване и противодействие на антиромската реч на омразата в онлайн пространството. Българското изследване стъпва върху предходни европейски проучвания и има за цел да:

- Предостави цялостни данни за разпространението, въздействието, авторите и адресатите на антиромската реч на омразата, както и за реакциите на обществото и онлайн платформите.
- Повиши броя на докладваните случаи чрез разпространение на информация за наличните механизми за сигнализиране.
- Подкрепи разработването на политики и мерки за противодействие на онлайн дискриминацията, основани на доказателства.

1.3 Обучение и изграждане на капацитет

През септември 2024 г. във Варна се проведе четиридневно обучение Training for Trainers, посветено на възникването, ескалацията и противодействието на онлайн речта на омразата. От България в обучението участваха Ахмед Хасанов, Гамзе Максимова и Мария Христова, които впоследствие предадоха придобитите знания на Диян Данков - адвокат и правен консултант в Асоциация Интегро с дългогодишен опит в областта на правата на човека и антидискриминационната работа - и на Вергил Ибрям, активист с опит в национални изследвания на речта на омразата (2010-2014 г.).

В периода 26-29 август 2025 г. националните наблюдатели проведоха обучение за изграждане на капацитет с 22 участници.

Сесиите бяха насочени към идентифициране на основните причини за речта на омразата и осмисляне на собствените предразсъдъци; разбиране на тенденциите в разпространението на антиромската реч на омразата в онлайн среда; усвояване на инструментите за сигнализиране на проблемно съдържание; и анализ на основните тематични области на антиромската реч на омразата, свързани с жилищните условия, социално-икономическото положение, образованието и културните аспекти.

1.4 Законодателна и нормативна рамка

В България противодействието на речта на омразата и антиромския расизъм се урежда от редица национални закони и институционални механизми. Наказателният кодекс криминализира публичното подбуждане към насилие или дискриминация на етническа основа, а Законът за защита от дискриминация изрично признава антиромския расизъм като форма на расова дискриминация. Законът за защита на личните данни и Законът за електронните съобщения уреждат обработването на лична информация в дигитална среда. Комисията за защита от дискриминация (КЗД) разглежда сигнали и случаи, свързани с реч на омразата, въпреки ограничените си правомощия по отношение на модерирането на съдържание в частни онлайн платформи. Съветът за електронни медии (СЕМ) осъществява наблюдение и може да налага санкции при онлайн медийно съдържание, което подбужда към дискриминация или нарушава човешкото достойнство.

На европейско равнище Регламентът за цифровите услуги (Digital Services Act - DSA) и Директивата за борба с расизма и ксенофобията изискват от онлайн платформите да предприемат действия срещу дискриминационно и изпълнено с омраза съдържание. Актът за изкуствения интелект (AI Act, 2024 г.) въвежда изисквания за прозрачност и отчетност при автоматизираното разпространение на съдържание, като забранява използването на алгоритми, водещи до дискриминация или профилиране на основата на етническа принадлежност или религия. Тези европейски механизми допълват националното законодателство и създават обща правна рамка за противодействие на антиромската реч на омразата в онлайн среда.

Примери и тенденции в България (2023–2025 г.)

Медийни случаи и публични изявления:

- През периода 2023–2024 г. публични личности направиха изявления с антиромски внушения, които предизвикаха широк обществен отзвук в медиите и социалните мрежи. Това показва, че речта на омразата не е само онлайн явление, а е тясно свързана с публичния дискурс и политическата среда.
- **Националният доклад за мониторинг на речта на омразата**, изготвен от Асоциация Интегро въз основа на данни от проекта Peer Education to Counter Antigypsyist Hate Speech Online (PECAO), показва, че значителна част от публикациите съдържат стереотипи и обобщения (напр. „ромите винаги са бедни", „те не искат да се интегрират", „ромите разчитат единствено на социални помощи", „не плащат данъци и живеят на гърба на държавата"), които създават основа за нормализиране на дискриминацията.

  Активността в социалните мрежи се увеличава, като групи и страници с неясен произход разпространяват мемета, които осмиват ромите или използват обиден жаргон – тенденция, потвърдена и от анализа на Асоциация Интегро.

Статистически констатации от доклада на Асоциация Интегро:

- Около 60% от анализираните публикации, свързани с ромските общности, ги представят в негативен контекст, докато едва 15% ги отразяват в положителна или балансирана светлина.
- Коментарите в онлайн среда често разкриват двойни стандарти, например когато се критикува получаването на социални помощи от роми, без да се отчитат структурни фактори като историческата дискриминация или бариерите пред достъпа до образование.
- Платформите често реагират бавно или изобщо не реагират на съдържание, което нарушава принципите на медийната етика и правата на човека – констатация, която съвпада с изводите от анализи на европейско равнище относно модерирането на речта на омразата.

Законодателни и институционални рамки:

Национално равнище:

- **Наказателен кодекс –** криминализира подбуждането към насилие или омраза, основано на етническа принадлежност, религия или други защитени признаци.
- **Закон за защита от дискриминация –** определя и урежда механизмите за защита срещу дискриминация, включително на етническа основа, като сигналите и жалбите се разглеждат от Комисията за защита от дискриминация (КЗД).
- **Законодателство в областта на медиите и онлайн платформите –** създава регулаторна рамка за съдържанието, макар че по отношение на социалните мрежи продължават да съществуват значителни правни и технологични пропуски.

Европейско равнище:

- **Директива 2000/43/ЕО (Рамкова директива относно расизма и ксенофобията) –** изисква от държавите членки да криминализират определени форми на дискриминация и реч на омразата.
- **Регламентът за цифровите услуги (Digital Services Act – DSA) –** въвежда задължения за онлайн платформите, свързани с модерирането на съдържание, прозрачността и механизмите за подаване на жалби.
- **Актът за изкуствения интелект (AI Act)** – въвежда изисквания към алгоритмите с цел предотвратяване на дискриминационно профилиране въз основа на етническа принадлежност или други защитени признаци.

Тези правни рамки създават предпоставки за координирани действия между държавите, гражданското общество и онлайн платформите; въпреки това прилагането им продължава да среща трудности, както показват и констатациите от доклада на Асоциация Интегро.

Анализ на мерките за противодействие и препоръки:

- Модерирането и сигнализирането на съдържание остават недостатъчно развити: много потребители не знаят как да подадат сигнал, процедурите са бавни, а резултатите често са непредвидими.
- Обученията, проведени в рамките на ТААО, имат ключово значение за изграждането на капацитет сред активисти, юристи и млади хора, като им помагат да разпознават „меките" форми на реч на омразата (мемета, подигравки, ирония) и потенциала им да ескалират.

Препоръки:

- Да се засили сътрудничеството между Комисията за защита от дискриминация (КЗД), онлайн платформите и организациите на гражданското общество чрез дългосрочни партньорства и прилагане на механизми тип „trusted flagger" (доверени сигнализиращи организации).
- Да се адаптират обученията за учители, медиатори и младежки работници, така че да могат да разпознават и да реагират на прояви на антиромски расизъм в онлайн среда.
- Да се осигури систематично събиране и анализ на данни за случаи на реч на омразата, с цел планиране на целенасочени интервенции.
- Да се повиши прозрачността и отчетността на онлайн платформите чрез публикуване на публични доклади за модерирането на съдържание и чрез отговорно използване на автоматизирани инструменти в съответствие с изискванията на Регламента за изкуствения интелект (AI Act) и Регламента за цифровите услуги (Digital Services Act – DSA).

МЕТОДОЛОГИЯ

2.1 Цел и дизайн

Изследването, проведено в България в рамките на проекта ТААО, е разработено на основата на комбиниран качествен и количествен анализ на антиромските нагласи и онлайн речта на омразата. Основният акцент е поставен върху социалните мрежи - Facebook, TikTok, X, YouTube и Instagram - както и върху по-малки, но активни дигитални платформи.

Мониторингът обхвана периода октомври 2024 г. - септември 2025 г., като се прилагаше унифицирана европейска методология за идентифициране, проследяване и докладване на реч на омразата и антиромско съдържание. Основен инструмент за събиране на данни служеше специализираният Monitoring Tool (MT), използван от всички партньорски организации.

Основните цели на изследването включват:

- Идентифициране на явни и кодирани форми на антиромска реч на омразата.
- Анализ на публичния дискурс и негативните представяния на ромските общности в онлайн пространството.
- Проследяване на разпространението на негативни публикации, динамиката на коментарите и профилите на аудиторията.
- Оценка на въздействието на негативните коментари върху по-широката обществена среда.
- Наблюдение на реакциите на платформите след подадени сигнали за нарушения.

- Проследяване на броя и видовете премахнати публикации като индикатор за ефективността на модерирането.

Общо 528 публикации, съдържащи явни или кодирани прояви на антиромско съдържание, бяха идентифицирани и докладвани чрез специално създаден Google Form, паралелно с директно сигнализиране към съответните платформи. Екипът систематизира събраните данни в обобщена електронна таблица, което позволи проследяване на дела на сигнализираното съдържание, което впоследствие е било премахнато - важен количествен индикатор за ефективността на модерацията, особено предвид факта, че много процеси на модериране разчитат на алгоритми с ограничена точност при разпознаването на реч на омразата.

Събраната информация беше обобщена в четири тримесечни доклада, споделяни чрез онлайн платформа, достъпна за всички партньори. Данните бяха анализирани централизирано от унгарските партньори Roma Versitas, които изготвяха международни обобщения за всеки отчетен период.

Отчетни периоди:

- Октомври - декември 2024 г. (качен на 30.01.2025 г.)
- Януари - март 2025 г. (качен на 30.04.2025 г.)
- Април - юни 2025 г. (качен на 05.07.2025 г.)
- Юли - септември 2025 г. (качен на 28.10.2025 г.)

Опитът, натрупан в рамките на предходния проект РЕСАО, значително улесни процеса на работа. Всички партньорски екипи, включително Интегро, усъвършенстваха своята дейност чрез редовни срещи, екипни обсъждания и експертна обратна връзка от проф. Илеана Ротару. В отделни случаи се наблюдаваше припокриване на публикации, докладвани от различни наблюдатели - подобни несъответствия бяха своевременно идентифицирани и отстранявани в рамките на редовните екипни срещи, което допринесе за по-висока последователност и надеждност на данните.

2.2 Участници - Българският екип за мониторинг

Интегро сформира екип от петима млади роми, активни потребители на социалните мрежи, с разнообразен професионален

и образователен опит. Съставът на екипа беше умишлено изграден така, че да обхваща различни житейски и професионални перспективи: адвокат с опит в антидискриминационната работа, специалист от общинската администрация, ученици в гимназиален етап, студент и дългогодишен активист в областта на правата на ромите. Всички участници демонстрираха висока мотивация и ангажираност за разпознаване и противодействие на речта на омразата.

Профили на младежките наблюдатели (България)

Участник	Професионален и образователен профил	Предишен опит	Мотивация за участие
Диян Данков	36 г., адвокат, гр. Разград; женен, с две дъщери	15 години опит в Интегро; ръководител на екип по доброволчески и младежки програми	Дългогодишен интерес към правата на човека; желание да допринесе с професионалната си експертиза
Мария Христова	22 г., гр. Котел; главен специалист „Култура и социални дейности“; студент по английска филология	Участие в проекта PECAO и доброволчески мониторинг	Активна в онлайн пространството; стремеж да подкрепя реална промяна и по-ефективен контрол върху съдържание, съдържащо реч на омразата
Гамзе Максимова	19 г., гр. Разград; завършил ПМГ; проявява интерес към педагогиката	Участие в предходни проекти на Интегро	Мотивиран да противодейства на речта на омразата и да вдъхновява други млади роми
Ахмед Хасанов	17 г., гр. Разград; ученик в ПМГ, специалност „Приложно програмиране“	Участие в проекта SafeNet (2022–2024); сигнализиране за случаи на дискриминация	Желание за по-безопасна онлайн среда; чувство за солидарност и лична отговорност
Вергил Ибрям	36 г.; женен; дългогодишен активист и бивш координатор в Интегро; студент в РУ	Опит в мониторинга на медийни представяния и реч на омразата	Ангажираност към правата на ромите и към осигуряването на висококачествен мониторинг

Всички наблюдатели участваха в двуседмичен онлайн консултационен семинар, воден от проф. Ротару, насочен към идентифициране, категоризиране и докладване на реч на омразата в контекста на националното законодателство и методологията на проекта.

През септември 2024 г. част от екипа участва в четиридневен обучителен лагер, който допълнително укрепи капацитета им. Придобитите знания бяха споделени с други млади хора по време на националния обучителен семинар, проведен в периода 26-29 август 2025 г. във Варна.

Вергил Ибрям, като опитен член на Интегро, пое отговорността за вътрешната координация на екипа за мониторинг, като отчетността се осигуряваше чрез постоянна комуникация и ежемесечни онлайн срещи. Съвместно с координатора на Работен пакет 4, Камен Макавеев, те ръководеха процеса, предоставяха регулярна обратна връзка и при необходимост адаптираха методологията.

2.3 Инструменти и процедури на мониторинга

Изследването се основаваше на Monitoring Tool (MT) - структуриран аналитичен инструмент, адаптиран от методологията на проекта PECAO (2020-2022), позволяващ на наблюдателите да документират, анализират и оценяват публикации, съдържащи реч на омразата или антиромско съдържание.

На българския екип беше възложено да докладва минимум осем случая месечно. Всички случаи се регистрираха чрез Google Forms в съответствие с указанията в Приложение 1, като най-малко три от докладваните публикации се описваха подробно с цел осигуряване на качествен анализ.

Най-голям брой случаи бяха установени във Facebook, TikTok, YouTube и Instagram. В много от тях речта на омразата не се ограничаваше само до първоначалната публикация, а в значителна степен се проявяваше в коментарите под нея. Наблюдателите отбелязаха една особено показателна закономерност: дори когато новините бяха представени неутрално или включваха положително отразяване, наличието на визуално разпознаваеми роми често предизвикваше вълна от негативни, стереотипни и изпълнени с омраза коментари. Тази тенденция беше потвърдена и от международните партньори, което

подчертава дълбоко вкоренените предразсъдъци, изразявани нерядко от хора с малък или никакъв пряк контакт с ромските общности.

Инструментът за мониторинг включва шест аналитични раздела:

1. Основна информация - автор, адресат, платформа, достъпност
2. Основна тема - теми и фактори, провокиращи публикацията
3. Тип и стил на съдържанието - текстово, визуално, емоционален тон
4. Интензивност и ниво на омразата - явна или кодирана
5. Мерки за реакция - подаване на сигнали, противодействие, отговори
6. Лични наблюдения и последващи действия - впечатления на наблюдателя, реакция на платформата

След обработката на данните екипът колективно избираше случаите за включване в тримесечните обобщени доклади. Когато една и съща публикация се появяваше на няколко платформи, се избираше версията с най-широко разпространение. Всички данни бяха систематично интегрирани в националните обобщени доклади, формиращи част от консолидираната европейска база данни, координирана от Amaro Drom e.V.

РЕЗУЛТАТИ

3.1 Количествен компонент на изследването

3.1.1 Резюме

Количественият компонент на изследването предоставя ключова информация за обема и характеристиките на анализираните публикации, съдържащи антиромска реч на омразата. През периода на мониторинга българският екип идентифицира и анализира общо 528 случая, отговарящи на критериите за насоченост срещу ромската общност и обхващащи различни форми на реч на омразата. Категоризацията бе извършена чрез Monitoring Tool, разработен въз основа на методологията на доц. д-р Илеана Ротару. Анализът е структуриран около шест аналитични измерения: обща информация; основна тема и представяне; тип и стил на съдържанието; нетекстови елементи и визуални материали; интензивност и нива на омразата; и противодействия и емоционални реакции.

Обработката на докладите разкри важна системна закономерност. Първоначалните сигнали обикновено се преглеждат от автоматизирани AI системи, като човешки модератор се намесва едва след подаване на изрично искане за преоценка. Тази двустепенна система забавя премахването на публикации с явна реч на омразата и създава условия за по-широко разпространение на проблемното съдържание.

Значителен брой публикации с антиромска реч на омразата продължават да циркулират свободно, достигайки до широка и нерядко уязвима аудитория. Последствията са особено тревожни по отношение на младите хора, включително младите роми, които

могат да изпитват страх, тревожност и несигурност. В някои случаи негативните стереотипи се усвояват вътрешно, което може да доведе до самоналагащи се ограничения или адаптиране към изкривените обществени представи за ромската общност.

Тези наблюдения подчертават системния характер на проблема и необходимостта от целенасочени мерки на няколко нива едновременно - по-строги регулации, по-ефективно прилагане на съществуващите механизми за защита и по-активна роля на платформите в предотвратяването на разпространението на токсично съдържание. Отговорността е споделена между институциите, законодателите и самите платформи.

3.1.2 Количествени анализи

А. Обща информация – общ брой докладвани случаи и социални мрежи: типове акаунти и профили

А.1 Общ брой идентифицирани случаи и социални мрежи

Общият брой документирани случаи през периода на мониторинга възлиза на 528 публикации, разпределени по пет основни социални мрежи, както следва: Facebook – 332 случая (62,88%); TikTok – 102 случая (19,32%); Twitter (X) – 38 случая (7,2%); YouTube – 29 случая (5,49%); Instagram – 21 случая (3,98%); и Други медии – 6 случая (1,14%).

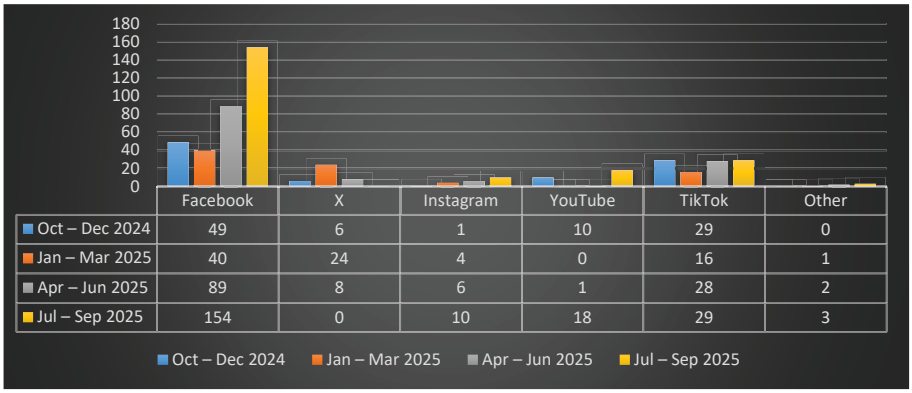
Таблица 1. *Брой публикации по тримесечия*

Социална мрежа	Октомври - декември 2024	Януари – март 2025	Април – юни 2025	Юли – септември 2025	Общо
Facebook	49	40	89	154	**332**
X	6	24	8	0	**38**
Instagram	1	4	6	10	**21**
YouTube	10	0	1	18	**29**
TikTok	29	16	28	29	**102**
Other	0	1	2	3	**6**

Диаграма 1 визуализира разпределението на антиромската реч на омразата по социални мрежи, като ясно показва, че Facebook е доминиращата платформа с 62,88% от всички документирани случаи, следвана от TikTok (19,32%) и YouTube (5,49%).

Бележка: *По време на мониторинга екипът реши колективно да не докладва случаи, по-стари от шест месеца. Реалният брой антиромски публикации, циркулиращи в интернет, вероятно е значително по-висок от представените тук данни.*

Диаграма 1



	Facebook	X	Instagram	YouTube	TikTok	Other
Oct – Dec 2024	49	6	1	10	29	0
Jan – Mar 2025	40	24	4	0	16	1
Apr – Jun 2025	89	8	6	1	28	2
Jul – Sep 2025	154	0	10	18	29	3

Анализът показва, че най-голям дял от случаите на антиромска реч на омразата са разпространени в социалната мрежа Facebook (62,88%), следвана от TikTok (19,32%) и YouTube (5,49%).

A.2 Типове акаунти и профили, разпространяващи антиромска реч на омразата

Освен анализа на платформите, изследването се фокусира и върху профилите на потребителите, които най-активно публикуват и споделят антиромско съдържание. Проследяването на авторите на публикациите позволява да се картографират основните разпространители на реч на омразата и да се измери интензивността на публикуване по тип акаунт.

Таблица 2: *Основни профили, разпространяващи онлайн реч на омразата*

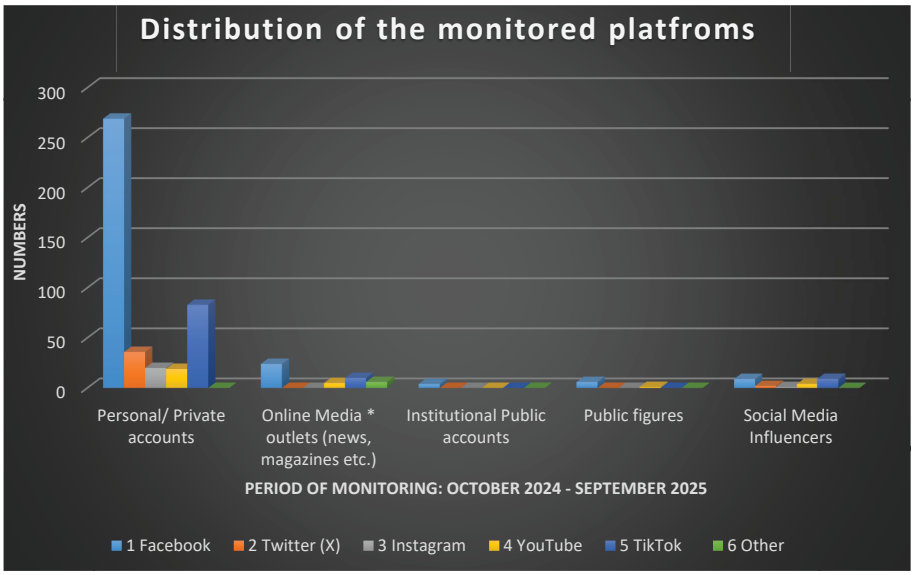
Социална медия		Тип на акаунта/профила				
		Лични/ частни профили	Онлайн медийни източници (новинарски сайтове, списания и др.)	Публични институционални акаунти	Публични личности	Инфлуенсъри в социалните мрежи
a		b	c	d	e	f
1	**Facebook**	269	24	4	6	9
2	**Twitter (X)**	36	0	0	0	2
3	**Instagram**	20	0	0	0	1
4	**YouTube**	19	5	0	1	4
5	**TikTok**	83	10	0	0	9
6	**Other**	0	6	0	0	0

Анализът на събраните данни разкрива, че личните акаунти на отделни потребители представляват основния източник на антиромска реч на омразата онлайн – 427 публикации от индивидуални профили, или около 81% от всички документирани случаи. Този висок дял подчертава, че разпространението на такова съдържание се движи предимно от масовото участие на отделни потребители, а не от организирани групи или институции.

Вторият по значимост източник са онлайн таблоиди и медийни източници с официални профили в социалните мрежи – 45 публикации (8,52%), разпространявани основно чрез Facebook, YouTube и TikTok. Макар делът им да е по-нисък, медийните публикации носят по-голям потенциал за разпространение и могат да увеличат видимостта и възприеманата легитимност на антиромските послания.

Диаграма 2 визуализира разпределението на документираните публикации според тип акаунт и социална мрежа. Индивидуалните акаунти доминират във всички платформи, докато медийните профили допринасят за усилване и по-широко разпространение на тези наративи.

Диаграма 2



B. Основна тема и представяне

Въз основа на методологията за мониторинг, използвана в проекта ТААО, бяха идентифицирани десет тематични категории, в които се разпределя антиромското онлайн съдържание:

1. Престъпления, извършени от роми
2. Социални аспекти — жилищни условия, социални помощи, бедност, миграция и др.
3. Образователни аспекти — отпадане от училище, условия на обучение, стипендии и др.
4. Здравни и санитарни аспекти — пандемии, достъп до болници, аборти и др.
5. Социални движения и НПО — протести, граждански права, представителство
6. Политика — политическо представителство, партии, избори
7. Ромски лидери, включително жени
8. Културни събития — музика, филми, театър и др.
9. Спортни събития — състезания, игри
10. Други теми, свързани с ромите

Таблица 3 представя резултатите от анализа на основните теми в 528 публикации. Тъй като всяка публикация може да попада едновременно в повече от една категория, общата сума надвишава броя на публикациите.

Таблица 3

Основна тема		**Позитивно +**	**Негативно 0**	**Неутрално -**
Престъпления, извършени от ромите	**a**	7	65	148
Социални аспекти – жилищни условия, социални помощи, бедност, миграция и др.	**b**	8	92	108
Образователни аспекти – отпадане от училище, условия на обучение, стипендии и др.	**c**	3	103	24
Здравни/санитарни аспекти – пандемии, достъп до болници, аборти и др.	**d**	1	106	28
Социални движения и НПО – протести, граждански права, представителство	**e**	4	107	19
Политика – политическо представителство, партии, избори	**f**	1	113	47
Ромски лидери – включително жени	**g**	0	113	0
Културни събития – музика, филми, театър и др.	**h**	10	68	12
Спортни събития – състезания, игри	**I**	0	148	2
Други – теми, свързани с ромите	**j**	10	91	126

Диаграма 3 илюстрира количественото разпределение на тематичните категории, а Схема 1 представя процентното разпределение на публикациите по тон - положителен, отрицателен и неутрален.

Аналитичните резултати очертават многопластов профил на онлайн дискурса около ромската общност и разкриват структурни пропуски в механизмите за модериране на съдържание в дигиталните платформи.

Диаграма 3

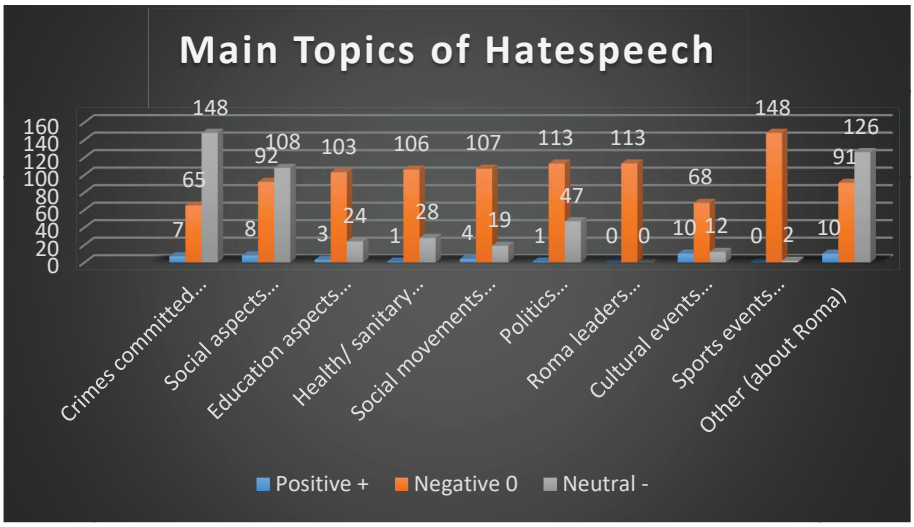
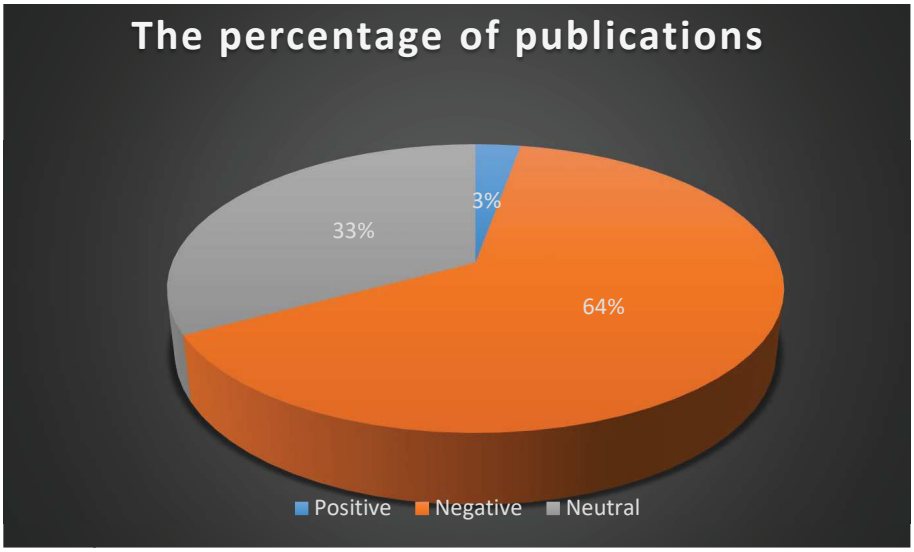


Схема 1 представя процентното разпределение на публикациите по тон в три аналитични категории - положителни, отрицателни и неутрални. Съотношението между трите категории надхвърля чисто описателното ниво и предлага аналитична перспектива за оценка на рисковете, свързани с разпространението на стереотипи, както и на възможностите за промяна на обществените нагласи чрез целенасочени интервенции.

Схема 1



Аналитичните резултати очертават сложен и многопластов профил на онлайн дискурса, свързан с ромската общност, като едновременно разкриват структурни пропуски в съществуващите механизми за модериране на съдържание в дигиталните платформи. Представените наблюдения се основават на систематичен мониторинг, при който оценките на екипа бяха калибрирани съгласно установен набор от инструменти за анализ на тон и реч на омразата.

Структурна доминация на отрицателния тон

Дялът на публикациите с отчетливо отрицателен тон (64%) представлява значителен индикатор за дълбоко вкоренени антиромски нагласи в онлайн средата. Тези публикации се характеризират с използването на дегуманизиращи категории, расови епитети и различни форми на явна или скрита враждебност.

Секциите с коментари под подобни публикации действат като „усилватели“ на токсичността, съдържайки устойчиви модели на екстремна вербална агресия, включително призиви към насилие. Това изисква критично внимание във връзка с националните и европейските стандарти за противодействие на расизма и речта на омразата.

Токсичност в контекста на формално „неутрално“ съдържание

Особено важно е да се отбележи, че дори публикации с положителен (3%) или неутрален (33%) тон предизвикват изразително негативни реакции. Това явление разкрива структурно вкоренен модел: съдържание, което само по себе си не нарушава етичните или професионални стандарти, може да функционира като тригер за масово генериране на стереотипни и дискриминационни коментари.

Проблемът се задълбочава, когато се споменава етнически произход – дори когато това е журналистически неуместно – което автоматично активира негативни колективни възприятия. Емпиричните данни изтъкват важен регулаторен въпрос: до каква степен формалното спазване на правилата на платформите е достатъчно, за да се предотвратят дискриминационни въздействия върху уязвими групи?

Ограничения на алгоритмичните механизми и регулаторни последици

Опитът в рамките на проекта ТААО ясно показва съществено несъответствие между мащаба на токсичните реакции и ефективността на алгоритмичните системи за откриване и премахване на проблемно съдържание.

Потребители с лични акаунти успешно заобикалят филтрите чрез графични варианти, кодирани обиди или неологизми, които алгоритмите не успяват да засекат. Това сочи към структурен дефицит в архитектурата на платформите: системите за модериране се фокусират предимно върху съдържанието на публикациите, без да анализират динамиката на реакциите в секциите за коментари.

От регулаторна гледна точка това изисква разработване на стандарти, които да обхващат не само основното съдържание, но и неговите социални последици и вторични ефекти.

Влияние върху публичния дискурс и политическата култура

Идентифицираните тенденции имат значими последствия за формирането на обществените възприятия. За много онлайн потре-

бители образът на ромската общност се конструира почти изцяло чрез съдържание с негативен тон или чрез токсичната динамика на коментарите. Това подпомага възпроизвеждането на социални предразсъдъци и може да улесни политическа радикализация, особено в условия на социална напрегнатост.

От гледна точка на европейските политики за дигитална демокрация, тези констатации подчертават трудността съществуващите регулаторни инструменти да защитят адекватно уязвимите групи, без да ограничават свободата на изразяване. В същото време те акцентират върху необходимостта от засилено сътрудничество между националните власти, платформите и гражданското общество.

C. Тип и стил на съдържанието

Анализът на типовете и стила на съдържанието предоставя ключова информация за начина, по който образът на ромската общност се представя в социалните мрежи и други онлайн канали. Този подход позволява не само количествено картографиране на публикациите, но и качествено разбиране на комуникационните практики, които формират обществените нагласи към ромската общност.

1) Тип на съдържанието

Публикациите се класифицират според типа на съдържанието и съответния тон, записан в инструмента за анализ - положителен, отрицателен или неутрален. Основните категории са:

- **Медийна информация** - репортажи, новинарски статии, анализи
- **Покани** - събития с културен, спортен или образователен характер, включително концерти и уебинари
- **Съобщения** - информация за ежедневни дейности, прессъобщения и административни уведомления
- **Авторски мнения и позиции** - редакционни материали, лични мнения и коментари
- **Реклами и препоръки** - маркетингови публикации, промоции, обяви за работа и пътувания

- **Любопитни случаи** - необичайни или специални събития, истории и хуманитарно съдържание
- **Развлекателно съдържание** - музика, видеа, филми и друго развлекателно съдържание
- **Други** - посочени подробно чрез инструмента за медиен мониторинг

Тази категоризация позволява идентифициране на доминиращите формати на онлайн съдържание, които влияят върху общественото мнение и репутацията на ромската общност.

2) Стил на съдържанието

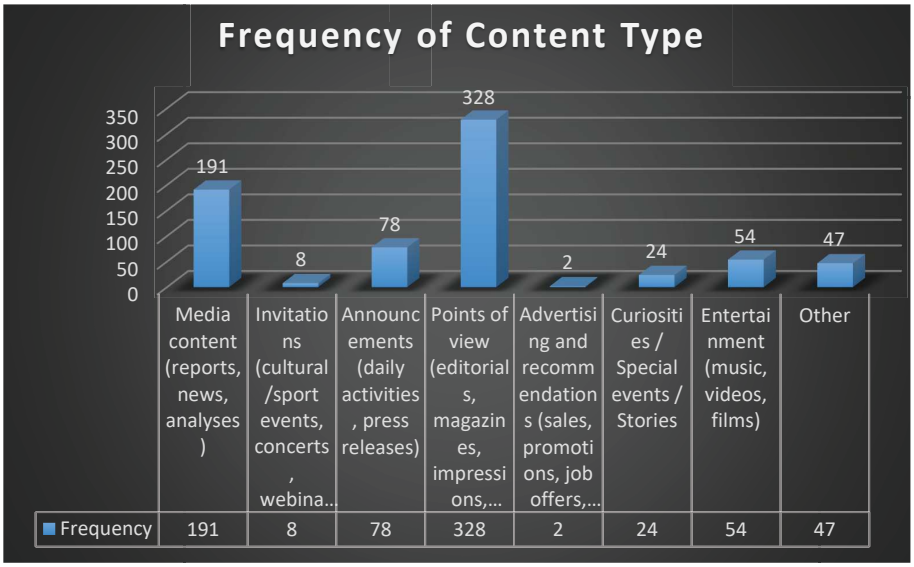
Съдържанието се класифицира и според комуникационния стил, отново разграничен по тон - положителен, отрицателен или неутрален. Основните стилове са разгледани подробно в раздел С.2, където е представен и графичният анализ на тяхното разпределение. Това разграничение позволява анализ на комуникационните стратегии, използвани онлайн, както и на потенциалното им въздействие върху формирането на стереотипи или положителни нагласи.

С.1 Анализ на типовете съдържание

Данните, представени в Диаграма 4, обхващат целия период на мониторинга и предоставят систематичен преглед на разпределението на различните типологии на съдържание, както и на техния принос за формирането на онлайн дискурса, включително антиромската реч на омразата.

Бележка: Посочените числа в таблиците и диаграмите (например 328, 191 и др.) не отразяват абсолютния брой публикации, тъй като общият обем на анализираните публикации е 528. Всяка публикация може да попада в повече от един тип или стил на съдържание - например една новинарска статия може едновременно да съдържа личното мнение на автора. Поради това тези стойности се представят като „честота на наблюдаваната типология", показваща колко пъти дадена типология е идентифицирана в целия набор от публикации.

Диаграма 4: *Честота на типовете съдържание*



Публикациите с авторски мнения и позиции са най-често срещаният тип съдържание, с честота 328, което представлява 44,8% от всички документирани случаи. Това показва, че значителна част от антиромската реч на омразата се разпространява чрез лични публикации на потребители с частни акаунти в социалните мрежи. Системният анализ на тези публикации подчертава ключовата роля на индивидуалната активност за формирането на общественото мнение.

На второ място по честота са медийните публикации, включително новинарски репортажи и материали, насочени към повишаване на обществената осведоменост, с 191 случая (26,09%). Това са най-често публикации от журналистически профили, официални акаунти на публични личности и инфлуенсъри, което демонстрира значението на традиционните и полупубличните канали за формирането на онлайн дискурса.

Трети са съобщенията, предназначени да информират участниците в различни социални групи (локални или тематични), с честота 78 случая (10,65%). Тези публикации се споделят основно в отворени или затворени групи във Facebook или Instagram и са насочени към потребители с конкретни интереси.

На четвърто място е развлекателното съдържание, включително антиромска реч в развлекателен контекст, с 54 случая (7,38%).

Тези публикации често заобикалят алгоритмичните системи за откриване поради сложността при разпознаване на фина или маскирана реч на омразата.

Останалите категории се разпределят, както следва:

- Други - 47 случая (6,42%)
- Любопитни факти - 24 случая (3,2%)
- Покани за събития - 8 случая (1,09%)
- Реклами и препоръки - 2 случая (0,27%)

Ключови изводи

Разпределението на честотата на типовете съдържание разкрива важна закономерност: антиромската реч на омразата не е концентрирана в един формат, а прониква в почти всички видове онлайн комуникация - от личните мнения и медийните репортажи до развлекателното съдържание и груповите съобщения. Това я прави особено трудна за системно идентифициране и ограничаване.

Доминирането на авторските мнения и позиции (44,8%) подчертава, че индивидуалната онлайн активност е основният двигател на антиромския дискурс. Когато обаче се вземе предвид и значителният дял на медийните публикации (26,09%), става ясно, че проблемът не се ограничава до анонимни потребители - той е вкоренен и в по-широките комуникационни практики на медии, инфлуенсъри и публични личности, чиито публикации имат несравнимо по-голям обхват и възприемана легитимност.

Особено тревожно е присъствието на антиромски послания в развлекателния формат. Именно там речта на омразата е най-трудно разпознаваема - маскирана като хумор, ирония или забавление - и следователно най-рядко засичана и модерирана от алгоритмичните системи на платформите. Това създава условия за нейното нормализиране и широко разпространение сред аудитории, които може и да не осъзнават дискриминационния характер на съдържанието, с което взаимодействат.

Тези наблюдения очертават необходимостта от многопластов подход при противодействието на антиромската реч на омразата - интервенции на нивото на дизайна на платформите, алгоритмичната

модерация, регулаторните рамки и образователните програми за дигитална грамотност.

Текстът е добър, но има два проблема: въвеждащият абзац е малко тромав, и последните два абзаца след списъка казват едно и също нещо с различни думи. Ето подобрената версия:

С.2 Графичен анализ на стиловете на съдържанието

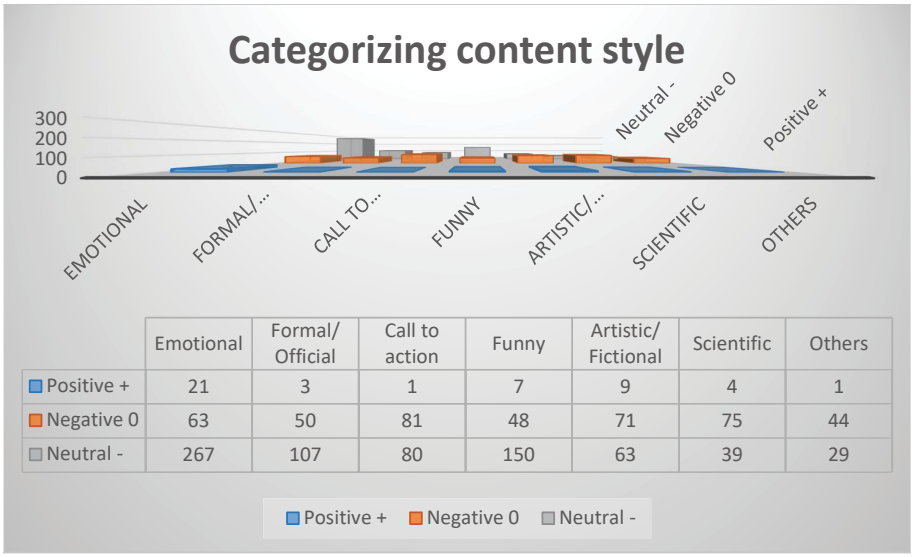
Графичният анализ на стиловете на съдържанието допълва категоризацията по тип, като поставя акцент върху начина, по който са оформени посланията, и върху потенциалното им емоционално или когнитивно въздействие върху аудиторията. Докато типът на съдържанието отговаря на въпроса „какво се публикува", стилът отговаря на въпроса „как се казва" - и именно това „как" до голяма степен определя дали дадено послание ще бъде възприето като неутрално или ще активира предразсъдъци.

Основните анализирани категории стилове включват:

1. Емоционален - съдържание, предназначено да предизвика силни емоции или емоционално ангажиране на аудиторията.
2. Формален/Официален - публикации, предаващи официални изявления, професионални съобщения или административна информация.
3. Призив за действие - съдържание, насърчаващо потребителите да предприемат конкретни действия, като участие в събития, споделяне на съдържание или подписване на петиции.
4. Развлекателен - мемета, видеа, хумористично или игриво съдържание, често включващо фини форми на антиромска реч.
5. Художествен/Фикционален/Фантастичен - творчески продукти, включително литературни, визуални или фикционални представяния на ромски личности или култура.
6. Научен/Информационен - съдържание, базирано на данни или изследвания, образователни публикации или фактически репортажи.
7. Други - всички останали стилове, идентифицирани чрез инструмента за мониторинг, непопадащи в основните категории.

Диаграма 5 показва относителната честота на всеки стил и позволява да се идентифицират кои комуникационни стратегии са най-разпространени в антиромския онлайн дискурс - кои генерират най-висока ангажираност, кои усилват стереотипи и кои влияят най-силно върху общественото възприятие. Тези данни имат пряко практическо значение за разработването на по-ефективни политики за модериране, за регулаторните рамки на платформите и за програмите за дигитална грамотност.

Диаграма 5



	Emotional	Formal/ Official	Call to action	Funny	Artistic/ Fictional	Scientific	Others
Positive +	21	3	1	7	9	4	1
Negative 0	63	50	81	48	71	75	44
Neutral -	267	107	80	150	63	39	29

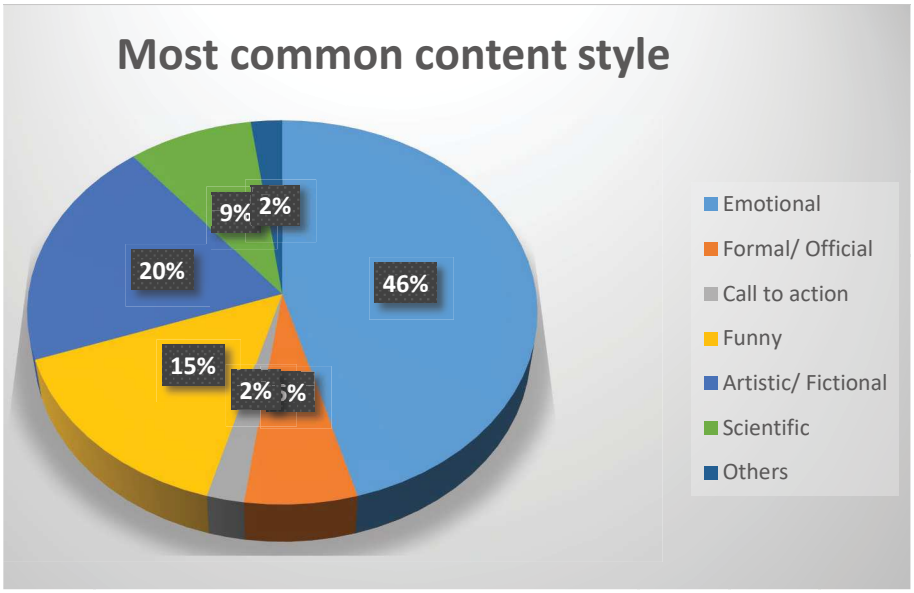
Схема 2 илюстрира разпределението на стиловете на антиромското съдържание в онлайн средата.

Най-често срещаният стил е емоционалният, който представлява 46% от всички случаи. Този тип изразяване е насочен предимно към директни обиди и дегуманизация на ромските личности, като същевременно мобилизира подкрепата на не-ромски потребители. Тези потребители често подсилват негативните послания чрез реакции, коментари или споделяне на визуални елементи, което допълнително вкоренява предразсъдъците и стереотипите.

Анализът показва, че емоционалното съдържание играе ключова роля за усилване на антиромския дискурс онлайн, демонстри-

райки как афективното рамкиране може да увеличи ангажираността и да затвърди дискриминационните наративи.

Схема 2



Вторият по честота стил е художествената форма на атака (20%), при която ромски артисти, публични личности или културни събития, свързани с ромската общност, стават обект на негативни публикации. Този стил използва креативни изразни средства, но с цел дискредитиране и подкопаване на публичния образ на ромите. Особено тревожно е, че подобно съдържание често се разпространява в среди, в които изкуството и културата се възприемат като неутрална територия, което допълнително затруднява разпознаването му като реч на омразата.

Третият по разпространение е хумористичният стил (15%), проявяващ се чрез подигравки, иронични мемета, илюстрации и хумористични текстове. Въпреки че се представя като шеги, такова съдържание играе значителна роля в нормализирането на омразата и превръщането ѝ в социално приемлива. Хуморът функционира като защитен механизъм - авторите могат да отхвърлят критиката с аргумента, че „просто се шегуват", докато посланието продължава да циркулира и да укрепва стереотипите.

Тези различни стилове на антиромско изразяване в онлайн пространството отразяват многопластовия характер на речта на омразата. Доминирането на емоционалния стил показва, че стереотипният образ на ромите се усвоява на лично ниво от авторите на публикациите, които приписват колективна вина на цялата общност без доказателства. Тази динамика създава сериозни рискове - онлайн омразата се разпространява бързо, трудно се регулира и допринася за среда, в която всички роми се възприемат като хомогенна група, независимо от индивидуалното им поведение или реалните им обстоятелства.

D. Нетекстови елементи и визуални материали

Анализът на онлайн съдържание с антиромска насоченост изисква систематичен мониторинг не само на текстови публикации, но и на нетекстови формати, включително изображения, видеа, мемета и карикатури. Тези визуални материали често циркулират без съпроводителен текст, което ги прави по-трудни за автоматично разпознаване и модериране от онлайн платформите. Много социални мрежи все още нямат достатъчно развити алгоритми, способни да идентифицират дискриминационно съдържание във визуални формати, създавайки условия за широко разпространение на визуално кодирана омраза.

Особено важно наблюдение по време на мониторинга е, че премахването на такива публикации често изисква изцяло човешка намеса. В редица случаи първоначалният автоматизиран преглед заключава, че съдържанието не нарушава стандартите на платформата, което налага повторно сигнализиране и ръчна намеса на модераторите. Този процес подчертава ограниченията на изкуствения интелект при идентифицирането на визуални форми на дискриминация и необходимостта от комбинирани механизми, включващи човешки контрол.

Наблюдателите също отбелязаха значителен брой случаи, при които визуалното съдържание не съответства на текстовото описание. Такива публикации често произхождат от псевдомедийни източници, популистки страници или профили на крайнодесни параорганизации, които умишлено използват заблуждаващи изо-

бражения, за да създадат негативни конотации за ромската общност. Снимки или видеа, взети от външни източници, медийни архиви или несвързани събития, се използват за изграждане на изкривена и манипулативна интерпретация на реалността.

В други случаи дори положително съдържание, което акцентира върху инициативи или постижения на ромски личности, предизвиква вълна от негативни реакции, провокирани единствено от визуалното изобразяване на ромски квартали или лица на роми. Това налага сигнализиране на тези публикации поради обидни и агресивни коментари под тях, които често ескалират до реч на омразата. Подобни реакции показват дълбоко вкоренени стереотипи, активирани дори от визуални маркери, които по същество не са негативни.

Поради тези причини методологията на мониторинга включва и категоризация на визуалните материали според тяхното въздействие върху аудиторията - положително, неутрално или отрицателно.

Ключови количествени констатации

През дванадесетмесечния период на мониторинг бяха анализирани 528 публикации с явна или скрита антиромска насоченост. От тях 409 съдържаха визуален елемент, което представлява 77,46% от всички публикации - дял, който сам по себе си е показателен за централната роля на визуалното съдържание в разпространението на антиромския дискурс онлайн.

Според наблюденията на екипа публикациите с визуално съдържание привличат значително повече внимание от аудиторията, включително и от ромски лица, които често реагират интензивно. Това разкрива тревожен процес на вътрешно усвояване на стереотипи - негативните визуални сигнали не само достигат до общността отвън, но биват усвоявани и възпроизвеждани от самите нейни членове. Когато стереотипите се интернализират по този начин, те могат да подкопаят самочувствието, да ограничат социалните амбиции и да нормализират дискриминацията като нещо „очаквано“ или „естествено“.

Таблица 6: *Брой визуални елементи и тяхната категория на въздействие*

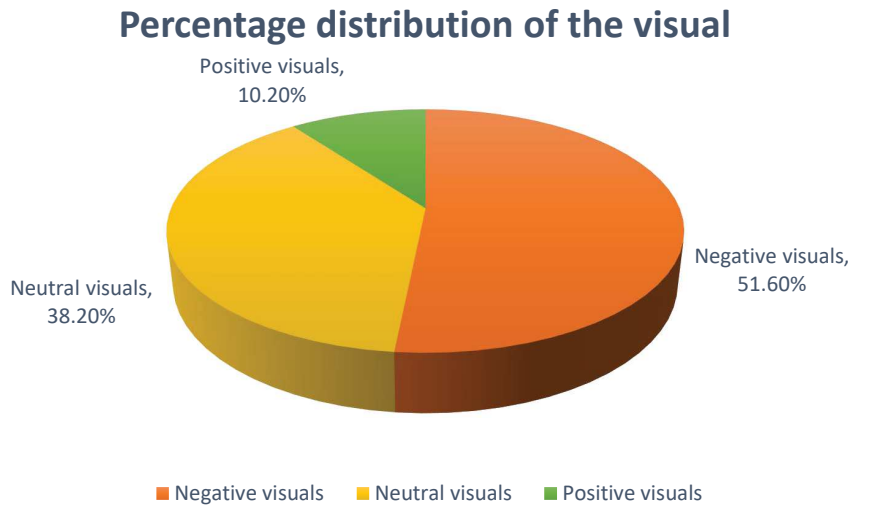
Нетекстова форма	Позитивно +	Негативно 0	Неутрално -
Снимки	8	100	52
Мемета/гифове	1	11	12
Карикатури	0	0	0
Мултимедийни материали (Reels, Stories)	4	16	8
Видеа	29	74	81
Анимации	0	10	3

Таблицата показва разпределението на визуалните елементи според тона:

- **Отрицателни визуални елементи – 211 броя (51,6%), представляващи най-голямата група.**
- **Неутрални визуални елементи – 156 броя (38,2%).**
- **Положителни визуални елементи – 42 броя (10,2%), най-малката категория.**

Това ясно показва, че отрицателните изображения доминират над останалите категории.

Схема 3



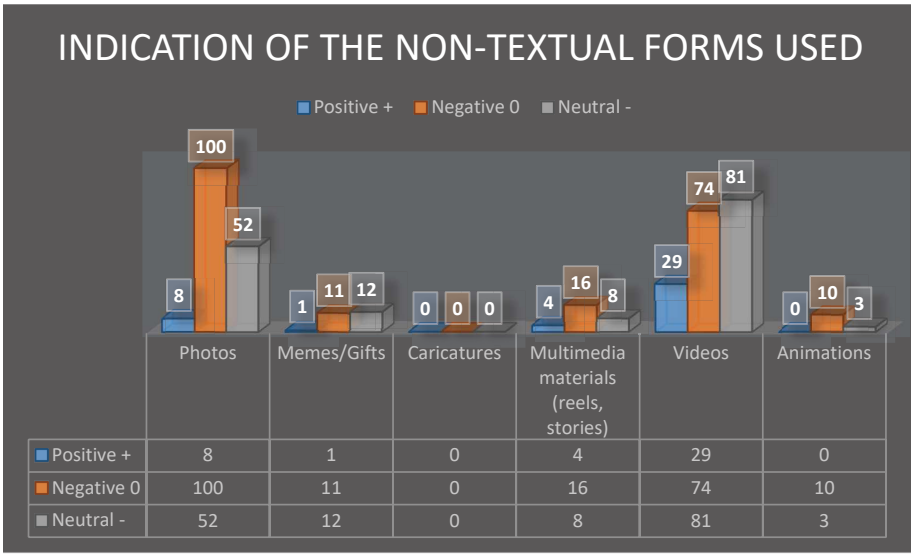
Според наблюденията на мониторите дори неутрални или положителни визуални материали често предизвикват негативни реакции, когато съдържат изображения, свързани с ромската общност. Това показва, че визуалните стереотипи са устойчиви и могат да активират враждебност независимо от действителното съдържание или послание на публикацията – достатъчно е самото присъствие на разпознаваем ромски контекст.

Негативните реакции и коментари под такива публикации допълнително затвърждават вредните стереотипи, създавайки условия за пренасяне на онлайн омразата в офлайн контексти. Този процес вече се проявява в нарастващ брой случаи на престъпления от омраза срещу роми през последните години – тревожен показател за реалното социално въздействие на онлайн дискурса, което далеч надхвърля виртуалното пространство.

Съвкупността от тези наблюдения очертава визуалните материали като един от най-трудно контролируемите, но и най-въздействащи вектори на антиромския онлайн дискурс. Алгоритмите за модериране все още не са достатъчно развити, за да ги разпознават надеждно, докато тяхното влияние върху обществените нагласи, затвърждаването на стереотипите и провокирането на агресивно поведение е добре документирано от мониторинга.

Следната диаграма илюстрира разпределението на типовете визуални елементи, които най-силно допринасят за разпространението на антиромския дискурс в онлайн пространството.

Диаграма 6



Нетекстовите и визуални материали се оказват ключов двигател на антиромския дискурс онлайн. Те не само се разпространяват по-лесно и остават по-трудни за автоматично разпознаване, но също така предизвикват силни реакции от аудиторията. Данните от мониторинга показват, че отрицателните визуални сигнали доминират, а дори неутрални или положителни изображения, свързани с ромската общност, често предизвикват враждебни коментари - явление, което разкрива дълбоко вкоренени визуални стереотипи и допринася за ескалацията на онлайн омразата. Разбирането на тези механизми е предпоставка за следващия елемент на анализа - интензивността и нивото на самата омраза.

Е. Интензивност и ниво на омразата - екология и степен на токсичност

Този раздел се фокусира върху степента на враждебност и риторическите форми, чрез които антиромската реч на омразата се проявява в наблюдаваните онлайн пространства. Методологията на изследването предвижда систематичен мониторинг на два основни типа онлайн омраза - явна и латентна, като инструментът за

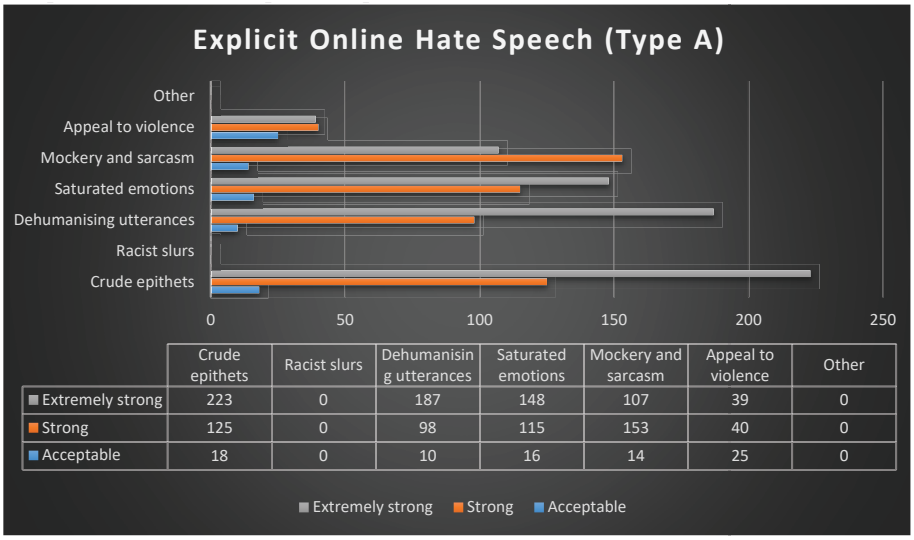
мониторинг ги различава чрез оценка на нивото на агресивност и лингвистичните стратегии, използвани за прикриване или усилване на дискриминацията.

Явна омраза (Тип А) - изразена чрез директни и лесно разпознаваеми лингвистични средства, включително:

- Груби обиди и расови епитети
- Дегуманизиращи изрази
- Наситени емоции (гняв, възмущение, враждебност)
- Подигравки и сарказъм
- Призиви към насилие

Този тип съдържание е непосредствено разпознаваемо и е насочено към открито дискредитиране на ромската общност.

Диаграма 7: *Явна онлайн реч на омразата (Тип А)*



	Crude epithets	Racist slurs	Dehumanising utterances	Saturated emotions	Mockery and sarcasm	Appeal to violence	Other
Extremely strong	223	0	187	148	107	39	0
Strong	125	0	98	115	153	40	0
Acceptable	18	0	10	16	14	25	0

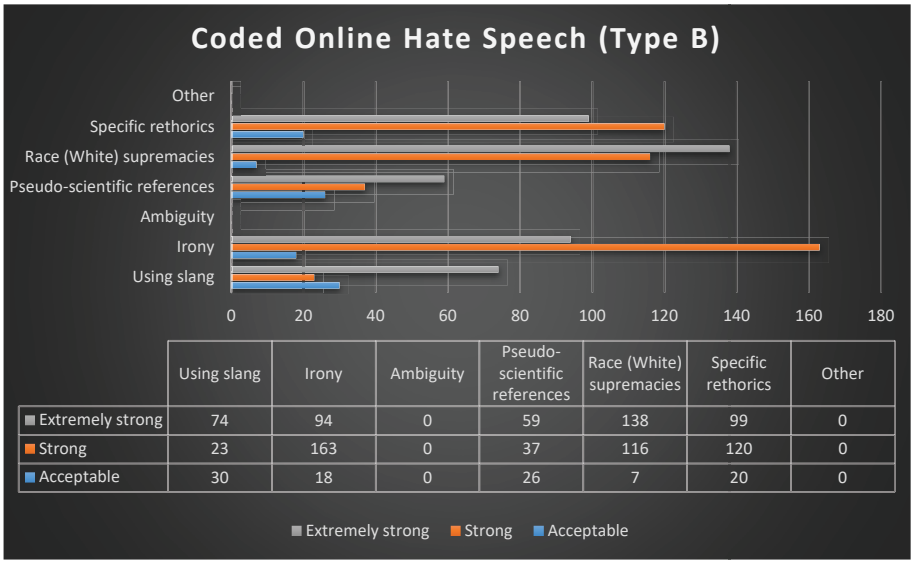
Латентна омраза (Тип В) - кодирана чрез специфични метонимични конструкции и изрази, включително:

- Използване на жаргон
- Ирония
- Двусмисленост
- Псевдонаучни позовавания
- Внушения за расово превъзходство

- Специфични риторични похвати
- Други

Този тип омраза е по-трудна за разпознаване и често заобикаля автоматизираните системи за модериране.

Диаграма 8: *Кодирана / латентна онлайн реч на омразата (Тип B)*



	Using slang	Irony	Ambiguity	Pseudo-scientific references	Race (White) supremacies	Specific rethorics	Other
Extremely strong	74	94	0	59	138	99	0
Strong	23	163	0	37	116	120	0
Acceptable	30	18	0	26	7	20	0

Това разграничение позволява по-точна оценка на интензивността и нивото на враждебност, както и на механизмите, чрез които онлайн съдържанието затвърждава стереотипи и стимулира агресивни нагласи към ромската общност.

Интензивността се измерва като броя на споменаванията на съответния тип омраза от наблюдателите в идентифицираните случаи. Освен това инструментът за мониторинг включва още една ключова характеристика на антиромското съдържание - нивото на омразата. То се определя въз основа на наблюденията и субективното възприятие на мониторите и се класифицира в три категории: приемливо, силно и изключително силно. Тези категории осигуряват основа за подробен анализ на наблюдаваните случаи, позволявайки идентифициране на публикациите с най-висока интензивност, както и на механизмите, чрез които антиромските послания се разпространяват и вкореняват онлайн.

Явна и латентна омраза - анализ на интензивността

Явната омраза се проявява най-силно чрез груби обиди, дегуманизиращи изрази и наситени емоции, като най-висока интензивност се наблюдава при публикации с изключително силен тон. Латентната омраза, от своя страна, се разпространява широко чрез иронични и псевдонаучни изрази, както и чрез косвени намеци за расово превъзходство, което значително усложнява автоматизираната модерация.

Количественият анализ показва, че при явната омраза (Тип А) публикациите с изключително силен тон съставляват 53,41% от всички докладвани случаи. Сред тях грубите обиди са най-чести (19,91%), следвани от дегуманизиращи изрази (14,18%), емоционално наситени изрази като гняв (11,22%) и подигравки (8,1%). Данните показват, че проявите с изключително силна явна омраза съставляват повече от половината от всички регистрирани случаи - красноречив показател за дълбочината на враждебността, циркулираща в онлайн пространството.

При латентната или кодирана омраза (Тип В) публикациите с изключително силен тон съставляват 45,3% от всички докладвани случаи. Най-честите прояви в тази категория включват специфични риторични похвати (15,66%), кодирани изрази на расово превъзходство (13,47%), ирония и иронични изрази (9,17%), както и жаргон и специфични думи, свързани с ромите, използвани за унижение на общността (7,22%). Макар делът на изключително силните случаи да е малко по-нисък в сравнение с явната омраза, те все пак представляват значителен обем с потенциално сериозно социално въздействие и способност да затвърдят стереотипите трайно.

Общият количествен анализ подчертава необходимостта от комбинирани стратегии за мониторинг и модериране, които интегрират автоматизирани инструменти с човешка експертиза. Резултатите показват, че както явната, така и латентната антиромска реч съдържат значителен дял публикации с изключително силен тон, изискващи прецизен и контекстуално чувствителен подход за идентификация и управление в онлайн средата.

Идентифицирани категории на токсично въздействие в онлайн средата

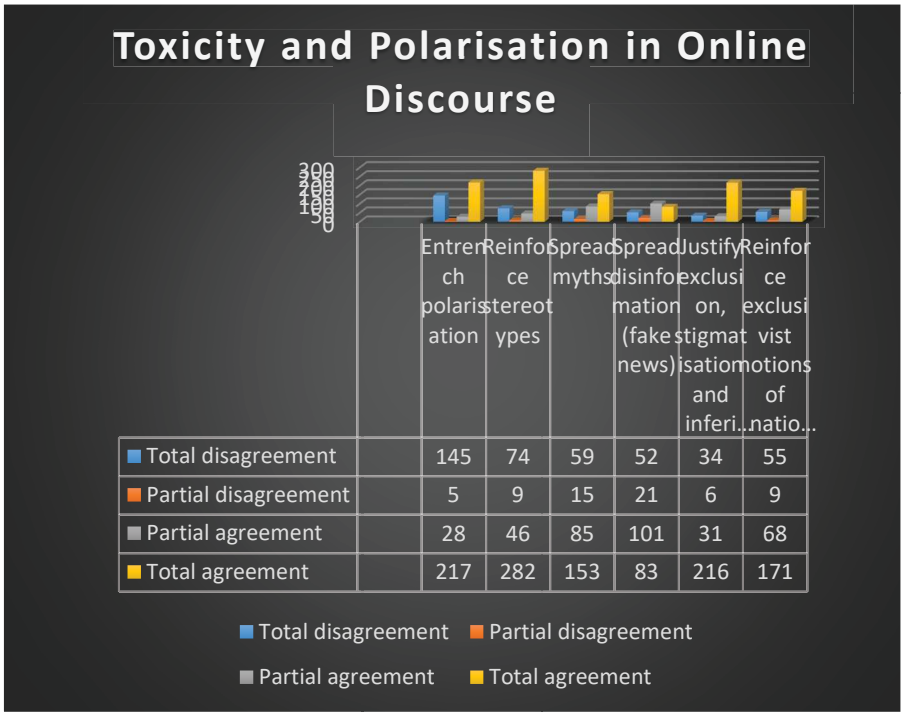
1. **Укрепване и потвърждаване на стереотипите - 282 споменавания** Това е най-често регистрираната категория. Наблюдават се устойчиви опити да се сведат ромите до колективни негативни характеристики - „престъпници", „мързеливи", „зависими от социални помощи". Такива опростени наративи заличават индивидуалността и възпроизвеждат стигмата в различни онлайн контексти - от коментари под новинарски статии до видеосъдържание в социалните мрежи.
2. **Задълбочаване на поляризацията - 217 споменавания** Тази категория включва публикации и дискусии, които противопоставят общности една на друга чрез разделителната логика „ние срещу тях". Поляризацията е особено разпространена в локални Facebook групи и политически дискусии, където антиромската реторика се вписва в по-широки популистки и националистически рамки. Тези онлайн „ехо-камери" усилват негативните нагласи и правят расизма социално приемлив.
3. **Оправдание на изключването, стигматизацията и маргинализацията - 216 споменавания** Наративите в тази категория прехвърлят отговорността за дискриминацията върху самите роми, представяйки ги като хора, които „не могат да се интегрират", „самонаказващи се" или „несъвместими с европейските ценности". Това отрича структурния аспект на неравенството и нормализира социалното изключване като нещо закономерно, а не като резултат от исторически и системни процеси.
4. **Укрепване на изключителни понятия за национална принадлежност и идентичност - 171 споменавания** Тази категория включва дискурси, които свързват „българска" или „европейска" идентичност с етническа хомогенност и представят ромите като „чужденци" или „заплаха" за националния характер. Тази стратегия не само поддържа антиромски настроения, но и разширява нетолерантността към други малцинствени и мигрантски групи, вграждайки антиромизма в по-широк ксенофобски наратив.
5. **Разпространение на митове - 153 споменавания** Типични примери включват преувеличени или напълно измислени твърдения за „престъпност", „висока раждаемост" или „приви-

легии“ на ромите. Тези митове поддържат стигматизацията и тревожно често остават неоспорени от другите потребители, което допълнително укрепва тяхната привидна достоверност.

6. **Разпространение на дезинформация (фалшиви новини) - 83 споменавания** Макар да е най-рядко отбелязваната категория, тя включва значими примери на манипулативно съдържание - псевдостатистически твърдения, фалшиви новинарски статии или конспиративни намеци. Разпространението им подчертава уязвимостта на широката аудитория към непроверена информация и лекотата, с която дезинформацията може да бъде поднесена като факт в дигиталната среда.

Диаграма 9 илюстрира динамиката на екологията на онлайн речта на омразата, насочена към ромската общност, като визуализира основните направления на този дискурс и най-честите токсични наративи, чрез които се формира стигматизацията (замразена срещу подвижна онлайн реч на омразата; Lentin, 2016).

Диаграма 9



F. Противодействия и емоционални реакции

Инструментът на българския екип за мониторинг събира данни за реакциите на платформите и мерките за противодействие, предприемани от младите наблюдатели след сигнализиране на антиромско съдържание. Този раздел представя както официалните действия на платформите, така и индивидуалните стратегии, използвани от младите участници при сблъсък с онлайн реч на омразата.

Анализът на данните показва, че най-често използваната реакция е сигнализиране на публикацията или нейния автор чрез системата на платформата (421 отговора „Определено да»), което отразява високото доверие на потребителите в съществуващите механизми и осведомеността им относно процедурите за докладване.

Други често прилагани стратегии включват:

- Оценка на аргументите и поведението на автора (155 „Определено да») - аналитичен подход, насочен към поставяне под съмнение на легитимността на наративите.
- Критика или разобличаване на тактиката на автора (87 „Определено да») - използва се при публикации с манипулативна реторика, стереотипи или дезинформация.
- Сериозен диалог или взаимодействие (258 „Определено да») - макар числото да изглежда високо, онлайн дебатите с автори на реч на омразата често се възприемат като емоционално изтощителни или неефективни, което обяснява смесеното отношение към тази стратегия.
- По-рядко използвани подходи включват:
- Взаимни реакции в същия тон (33 „Определено да») - отразяват ограничена готовност за пряк сблъсък.
- Подигравки и иронични реакции (5 „Определено да») - функционират като хумористичен начин за справяне с враждебността.
- Сигнализиране към институции или търсене на правна помощ чрез граждански организации (19 и 11 „Определено да») - относително рядко прилагана стратегия, което отразява ограничено доверие в ефективността на институционалните реакции.

Значителна част от наблюдателите изразиха несигурност относно някои стратегии (отговор „Не съм сигурен/на»), което подчертава неравномерното прилагане на практиките и необходимостта от допълнително обучение и подкрепа.

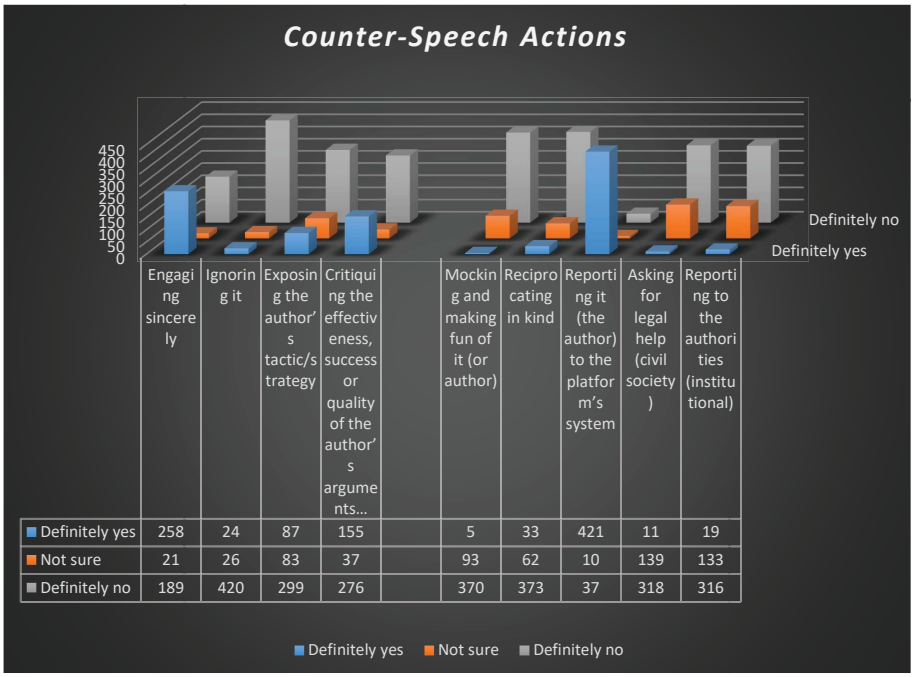
Тези резултати очертават няколко ключови наблюдения:

1. Сигнализирането към платформата остава най-достъпната и практически ефективна стратегия, осигуряваща видими резултати като временно блокиране, премахване на съдържание или предупреждения към авторите.
2. Индивидуалните и неформални подходи - критика, разобличаване на манипулативни стратегии или иронични реакции - имат потенциал за противодействие чрез реч и спомагат за развиване на аналитичните и емоционалните умения на наблюдателите.
3. Липсата на последователна обратна връзка от платформите създава несигурност относно ефективността на предприетите действия и подчертава необходимостта от по-голяма прозрачност и отчетност в дигиталното модериране.

В обобщение, данните от българския екип за мониторинг показват, че ефективното противодействие на онлайн омразата изисква едновременно институционални механизми и активно гражданско участие. Синергията между формалното сигнализиране към платформите и активния контрадискурс се очертава като ключов подход за намаляване на токсичността в онлайн средата - не като алтернативни варианти, а като взаимно допълващи се инструменти.

Диаграма 10 илюстрира разпределението на различните стратегии, използвани за противодействие на антиромското съдържание онлайн, въз основа на данни, събрани от българския екип за мониторинг.

Диаграма 10



	Engaging sincerely	Ignoring it	Exposing the author's tactic/strategy	Critiquing the effectiveness, success or quality of the author's arguments...		Mocking and making fun of it (or author)	Reciprocating in kind	Reporting it (the author) to the platform's system	Asking for legal help (civil society)	Reporting to the authorities (institutional)
Definitely yes	258	24	87	155		5	33	421	11	19
Not sure	21	26	83	37		93	62	10	139	133
Definitely no	189	420	299	276		370	373	37	318	316

Анализ на премахнатите публикации - допълнение към количествения анализ

Мониторингът на докладваните публикации изискваше не само количествено преброяване, но и внимателно проследяване дали съдържанието е действително премахнато или остава достъпно в публичното цифрово пространство. Тази стъпка служи като ключова връзка между количествения и качествения анализ, позволявайки по-точна оценка на ефективността на механизмите за модериране на платформите.

През периода на мониторинга бяха идентифицирани 528 публикации, съдържащи антиромско съдържание или реч на омразата, всички нарушаващи стандартите на платформите. От тях само 69 публикации бяха премахнати след сигнализиране от потребители, като по-голямата част са класифицирани под Типология А - явна и пряка реч на омразата.

Изразено като процент, едва 13,06% от докладваните случаи доведоха до реални действия от страна на платформите. Този про-

цент е алармиращо нисък, особено на фона на публично обявените политики за нулева толерантност към речта на омразата. Ограничената реактивност поражда основателни съмнения относно последователността и ефективността както на алгоритмичната, така и на човешката модерация, както и относно действителната ангажираност на платформите за защита на уязвими общности.

Тъй като националните наблюдатели докладваха само съдържание, публикувано през последните шест месеца, реалният обем на циркулиращата реч на омразата вероятно е значително по-голям от документираните 528 случая. В този контекст ниският процент на премахнати публикации предполага значително натрупване на проблемно съдържание, което остава достъпно, разпространява се свободно и постепенно се нормализира в цифровата екосистема.

Тази тенденция представлява системен риск. Онлайн средата се разширява непрекъснато, броят на потребителите нараства, като значителна част от тях са млади хора, на които често им липсва критична медийна грамотност за идентифициране на манипулативни или предубедени наративи. В резултат стереотипите и враждебните нагласи се усвояват като достоверни представи за ромската общност, въпреки че в повечето случаи са напълно откъснати от личен опит.

Като цяло тези констатации не само разкриват интензивността на антиромския дискурс онлайн, но и подчертават структурните пропуски в дигиталната модерация и неотложната необходимост от целенасочени интервенции - на институционално ниво, както и чрез граждански, образователни и медийни инициативи - с цел намаляване на устойчивостта и въздействието на онлайн омразата.

3.2 Качествен компонент на изследването

Качественият компонент допълва количествените резултати, представени в предходните раздели, като акцентира върху контекстуалните, емоционалните и интерпретативните измерения на антиромската реч на омразата, наблюдавани от българския екип за мониторинг. Освен числовите данни, разказите и наблюденията на мониторите разкриват сложността на циркулиращото онлайн съдържание с омраза, неговото емоционално въздействие върху участниците и възприеманите ограничения на съществуващите

механизми за сигнализиране и реакция. Следващите раздели представят ключовите изводи от качествения анализ, илюстриращи най-разпространените дискурси, модели на взаимодействие и емоционални реакции сред мониторите и аудиторията.

3.2.1 Основни тематични модели

Анализът на 528 документирани случая очертава сложен пейзаж от взаимно укрепващи се антиромски наративи. Тези наративи рядко съществуват изолирано - те се преплитат, създавайки среда, в която омразата се възпроизвежда, нормализира и оправдава.

Свързване на ромската идентичност с престъпност, „културен дефицит" и липса на образование

Най-устойчивият и най-често наблюдаван мотив беше редукционисткото представяне на ромската идентичност като неразривно свързана с престъпност, агресия или социална безотговорност. В значителен брой случаи се използваха генерализации като „те естествено крадат", „не искат да се образоват" или „така си са те". Тези твърдения, често рамкирани като „реализъм" или „здрав разум", не само заместват индивидуалната отговорност с етническа идентичност, но и затвърждават идеята, че социалното изключване е оправдано или неизбежно.

Емоционална поляризация и конструиране на морални йерархии

Наблюдаваните дискурси често изграждаха открито противопоставяне„ние срещу тях", представяйки ромите като морално, културно и социално „по-низши". Това морално рамкиране служеше за легитимиране на агресивни или наказателни нагласи, представяни като защита на„обществен ред", „патриотизъм" или„нормалност". Такова поляризиране засилваше възприятието за заплаха и допринасяше за повишени емоционални реакции, включително гняв, страх и отвращение.

Нормализиране на антиромските нагласи чрез хумор, ирония и „ежедневни шеги"

Друг широко наблюдаван модел включваше маскиране на враждебно съдържание като „шега", „забавен коментар" или „самоирония". Този привидно лек и игрив тон често скриваше дълбоко вкоренени стереотипи. Използването на хумор, особено под формата на мемета, позволяваше на потребителите да разпространяват омраза без отговорност, тъй като съдържанието лесно можеше да бъде отхвърлено като „несериозно" или „невинно". Резултатът обаче оставаше същият: затвърждаване на социалната дистанция и нормализиране на предразсъдъците.

Легитимиране на омразата чрез позоваване на свободата на словото

В значителен брой случаи омразата беше оправдавана чрез позоваване на „лично мнение", „сатира", „правото на критика" или „свободата на словото". Тази стратегия позволяваше на авторите да представят своите антиромски твърдения като част от легитимен публичен дебат, вместо като вредно съдържание. Мониторите наблюдаваха умишлено експлоатиране на двусмислия в правилата на платформите и националните регулаторни рамки, което позволяваше враждебните послания да останат онлайн и да се разпространяват без съществени последствия.

Преплитането на тези четири тематични модела създава среда, в която антиромските послания не просто циркулират, а се възпроизвеждат и затвърждават активно. Хумористичните форми на омраза служат като входна точка, нормализирайки негативните стереотипи и намалявайки обществената чувствителност към по-открито враждебни послания. Тази нормализация подготвя почвата за по-твърди наративи, като криминализиране на представите за ромите или моралната йерархизация „ние срещу тях". Позоваванията на свободата на словото от своя страна функционират като защитен механизъм, позволявайки на всички предходни форми да останат в публичното пространство, без авторите да бъдат държани отговорни или дори да осъзнават причинената вреда. По този начин

четирите модела се взаимно укрепват, допринасяйки за устойчива и трудно проникваща екосистема на онлайн антиромска омраза.

3.2.2 Анализ на избрани случаи

В рамките на мониторинговия процес бяха идентифицирани множество случаи на антиромска реч на омразата. За целите на този доклад се фокусираме само върху най-илюстративните и показателни примери, обхващащи както Тип А - явна реч на омразата, така и Тип В - латентна, косвена или кодирана агресивна реторика. Трябва да се подчертае, че представените примери са само минимална част от всички наблюдавани и докладвани случаи.

Случай 1:

Авторът на публикацията е известен бивш български журналист и настоящ член на Европейския парламент (MEP), печално из-

вестен с неподходящи изказвания и редовни обществени скандали. Засегнатият случай е още един пример за неговата проблематична реторика. Той се позиционира като против Еврозоната, еврото и европейските ценности, въпреки че изпълнява функциите си като МЕР. Освен това, той проявява открита враждебност към ромската общност.

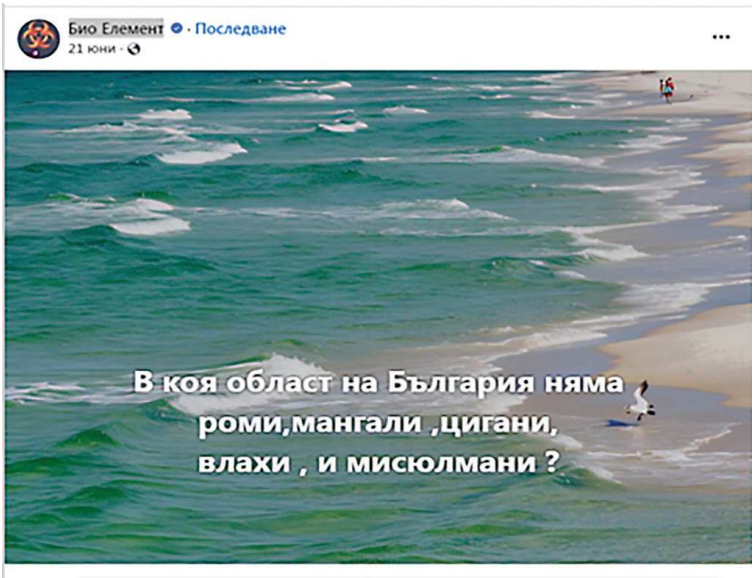
В скорошна публична публикация във Facebook той написа: *„На следващите избори… съдбата на България ще бъде хвърлена в един от два пътя — циганизация, корупция…“*

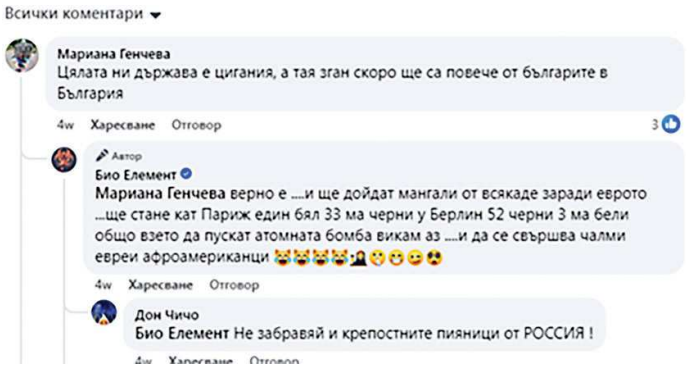
Това изявление е не само политически неприемливо, но и съдържа явна реч на омразата. Терминът „циганизация“ е расистки и ксенофобски израз, използван с цел да се внуши културен и социален упадък, който уж се дължи на присъствието на ромите. Публикацията разпространява вредни стереотипи и подхранва разделение, а използването й от избран политически представител я прави още по-тревожна.

Този пример представлява явна форма на реч на омразата. Въпреки че не съдържа визуални елементи, въздействието му е силно негативно. Публикацията е докладвана, но не е премахната от социалната мрежа, с аргумента, че „не нарушава стандартите на платформата“. Такова съдържание често се третира като „лично мнение“, създавайки широко сиво поле, което ограничава възможностите за ефективна модерация.

Допълнително притеснение представлява фактът, че съобщението е разпространено от публична личност и представител на България в Европейския парламент – аспект, който усилва легитимацията и въздействието на негативните стереотипи срещу ромската общност. Независимо от това, публикацията продължава да циркулира онлайн и предизвиква подкрепящи реакции, което подчертава устойчивостта и нормализирането на подобни нагласи в цифровата среда.

Случай 2:





Авторът на публикацията най-вероятно е администратор на публичната група „Bio Element“ - страница на инфлуенсър, която основно споделя видеа и снимки на спортни автомобили.

В публикацията авторът се обръща към аудиторията с въпроса: „Кой регион в България няма роми, цигани, „мангали“*, власи или мюсюлмани?”

Това не е просто случайно питане. Авторът се опитва да провокира реакции и развлечение сред читателите, но чрез употребата на унизителни и дискриминационни термини като „мангали“ и „цигани” допринася за допълнителна поляризация в обществото.

Тези термини нямат положителна конотация и се възприемат от ромската общност като дълбоко обидни.

Един от коментарите под публикацията гласи: „Цялата ни страна е една голяма циганска бъркотия и скоро тази сган ще изпревари българите в България."

В отговор авторът на оригиналната публикация написа: „Вярно е... и „мангали" ще идват отвсякъде заради еврото... ще стане като Париж - един бял човек, тридесет и трима чернокожи; в Берлин - петдесет и двама чернокожи, трима бели. В общи линии казвам, че просто трябва да пуснат атомната бомба и да свършат с това - тюрбани, евреи, афроамериканци."

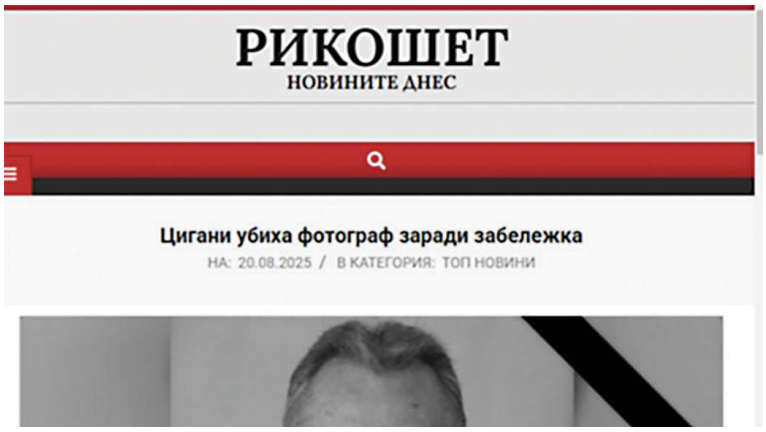
Този отговор съдържа изключително враждебен, расистки и насилствен език, включително:

- Употреба на расисткия обиден термин „мангали", насочен срещу ромската общност
- Расистки стереотипи и генерализации за чернокожи хора в европейските градове
- Пряк призив за масово насилие чрез предложение за използване на атомна бомба
- Изрази на омраза и дегуманизация, насочени към мюсюлмани („тюрбани"), евреи и афроамериканци

Случаят представлява ясен пример за реч на омразата и подбуждане към насилие - съдържание, което грубо нарушава стандартите на платформите и би трябвало да бъде незабавно премахнато. Показателен е и контекстът: публикацията произхожда не от анонимен профил, а от публична страница с широка аудитория, което многократно усилва нейния вреден потенциал.

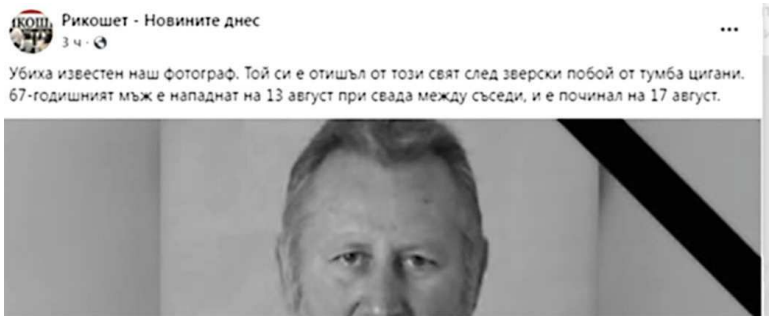
Бележка: „Мангали" е силно обиден расистки термин, използван срещу ромите в България.

Случай 3:
Официален уебсайт на таблоида Rikoshet, раздел „Top News“



Новинарският уебсайт съобщава за инцидент, при който известен местен фотограф е загубил живота си, след като е бил пребит от съседи заради жалба за силна музика. Съдържанието на статията като цяло спазва етичните журналистически стандарти, с изключение на заглавието: „Цигани убиха фотограф заради забележка".

Макар инцидентът да е изолиран случай, формулировката „Цигани убиха..." го представя като характерен за цялата ромска общност, като по този начин подхранва междуетническо напрежение и затвърждава вредни стереотипи. Посочването на етническата принадлежност на извършителите е журналистически неоправдано и не допринася за информирането на читателя – напротив, служи единствено за засилване на негативните асоциации. В придружаващата публикация във Facebook текстът допълнително задълбочава проблема: „Той (фотографът) почина след като беше брутално пребит от банда цигани."



Това е показателен пример за начина, по който медийното рамкиране може да превърне изолиран инцидент в обвинение срещу цяла общност. Само заглавието е достатъчно, за да предизвика негативни нагласи у читателя – дори без да се прочете самата статия. Именно тази способност да формира бързи и трайни впечатления прави подобното съдържание особено въздействащо и опасно от гледна точка на разпространението на антиромски нагласи.

Случай 4:

В скорошен епизод известен YouTuber и инфлуенсър, който разглежда различни теми и представя „факти", обяснява така наречената истинска същност на ромите, опирайки се на фалшиви доказателства, стереотипи и предразсъдъци, затвърждаващи негативния образ на ромската общност не само в България, но и в по-широк европейски контекст.

Видеото започва с подигравателно имитиране на ром, показан да танцува и да казва на снимача да не го записва. Оттук нататък цялото съдържание е изградено върху груби обобщения, представящи кражбата, джебчийството, нежеланието за учене и лошата хигиена като „най-добрите черти" на ромската общност - реторика, която

систематично заличава индивидуалността и свежда цяла общност до карикатурен образ.

Видеото е обидно и нарушава основните стандарти на платформата, но не е премахнато от YouTube. Причината е показателна: съдържанието генерира голям брой гледания и включва рекламни интеграции, което създава директен финансов стимул за платформата да не предприема действия. Това разкрива фундаментален конфликт на интереси - бизнес моделът на платформата се оказва в противоречие с нейните декларирани политики срещу речта на омразата.

Случаят е типичен пример за това как сарказмът и подигравката функционират като прикритие за дискриминационно съдържание. Хумористичната опаковка прави омразата по-лесно смилаема и по-трудно разпознаваема, а едновременно с това - по-широко споделяна и по-трудно модерирана от алгоритмичните системи на платформите.

3.2.3 Емоционални ефекти, онлайн динамика и ключови наблюдения от мониторинга на антиромската омраза в България

По време на мониторинга в рамките на проекта ТААО българският екип беше систематично изложен на голям обем антиромско съдържание, най-често във Facebook и TikTok. Речта на омразата на тези платформи се разпространяваше чрез подигравки, негативни стереотипи, мемета и кратки видеа, представяни като „хумор“ или „ежедневни наблюдения”. Този тип съдържание се оказа особено предизвикателен за интерпретация, тъй като враждебните послания често бяха маскирани като шеги, което улесняваше тяхната нормализация и осигуряваше висока ангажираност в платформите.

Въпреки предварителното обучение, повтарящото се излагане на такова съдържание причиняваше емоционално напрежение сред наблюдателите. Някои от тях съобщаваха за чувство на безпомощност и фрустрация, главно защото съдържанието рядко се премахваше, дори след многократни сигнали към платформите. Това затвърждаваше усещането, че алгоритмите и механизмите за модериране търпят или подценяват антиромския дискурс, особено когато той е представен в развлекателен формат. Емоционалното

натоварване на мониторите е важен показател сам по себе си - то разкрива колко интензивно и всепроникващо е присъствието на омразата в онлайн пространството дори за обучени и подготвени наблюдатели.

По време на мониторинга екипът идентифицира устойчиви онлайн екосистеми - страници, групи и профили, които систематично възпроизвеждат антиромски наративи под формата на псевдоновини, хумористични колажи, видеа с вирален потенциал или публикации, рамкирани като „граждански позиции". Концентрацията на такова съдържание във Facebook и TikTok се усилва допълнително от алгоритмите на платформите, които възнаграждават популярни и провокативни материали, без да отчитат тяхното дискриминационно въздействие.

Въпреки че мониторите обсъждаха потенциални реакции, екипът не предприе организирани контрадискурси или координирани действия - основният фокус остана върху наблюдение и докладване. Това решение беше методологически обосновано, но същевременно разкрива важно ограничение: дори систематичното документиране на речта на омразата не е достатъчно, ако не е съпроводено от активно противодействие. Саморегулацията на платформите очевидно не е в състояние да ограничи тези практики без външен натиск от страна на граждански организации, регулатори и самите потребители.

Събраните наблюдения подчертават устойчивото присъствие на антиромска омраза в българското онлайн пространство и необходимостта от по-ефективни механизми за модериране, по-ясни законови дефиниции на онлайн речта на омразата и по-систематична подкрепа за младите хора, ангажирани с мониторинг и противодействие на дискриминацията.

ДИСКУСИЯ

Националното изследване, проведено в рамките на проекта ТААО в България, разкрива устойчив и дълбоко вкоренен модел на антиромска враждебност в дигиталната среда, както и значителни разминавания между съществуващата законова рамка, политиките на платформите и реалния онлайн опит на потребителите. Количествените данни позволяват да се очертаят най-често срещаните форми и модели на враждебно съдържание, докато качествените наблюдения задълбочават разбирането на психологическите, социалните и алгоритмичните механизми, чрез които то функционира. Освен техническите ограничения на платформите, анализът подчертава и важното човешко и образователно измерение: предразсъдъци, стереотипи и липса на емпатия, които се усилват и нормализират онлайн.

4.1 Интерпретация на резултатите

Мониторингът потвърждава, че антиромските наративи са широко разпространени в българското онлайн пространство, особено във Facebook и TikTok, където те лесно се адаптират към текущите тенденции и алгоритмите на платформите. Освен явната реч на омразата (обиди, дегуманизация, призиви към насилие), значителна част от съдържанието се появява под формата на кодирани, иронични или псевдорационални изрази. Тези формати заобикалят както законовата класификация, така и автоматизираните филтри, маскирайки се като „хумор“, „лично мнение“ или „социален коментар”.

Алгоритмите на платформите допълнително усилват видимостта на подобни материали, като приоритизират съдържание, което предизвиква силни реакции. По този начин омразата се превръща едновременно в продукт и в инструмент за генериране на ангажираност - процес, който постепенно подкопава обществото доверие и социалната кохезия.

Особено тревожен е ниският дял на премахнатите публикации след сигнализиране - под 10%. Това поражда сериозни опасения относно ефективността и прозрачността на системите за модериране, доминирани от автоматизирани реакции с ограничен човешки контрол. Липсата на ясни обяснения както при премахване на съдържание, така и при отказ да се предприемат действия, допълнително подкопава доверието на потребителите и ги обезкуражава от докладване.

4.2 Значение и препоръки за заинтересованите страни

За онлайн платформите:

- Подобряване на точността на модерирането чрез комбиниране на AI системи с контекстуална човешка оценка.
- Повишаване на прозрачността на модераторските решения чрез ясни и конкретни обяснения за потребителите.
- Сътрудничество с ромски организации за разработване на обучителни материали и механизми за сигнализиране, съобразени с контекста.
- Въвеждане на по-подробни категории за докладване.
- Създаване на превантивни филтри за съдържание, съдържащо расистки символи, изрази или визуални модели.

За националните институции:

- Подобряване на координацията между Комисията за защита от дискриминация, Министерството на вътрешните работи, Министерството на електронното управление и други компетентни органи.
- Създаване на национална система за онлайн сигнализиране, която да позволява наблюдение на случаите в реално време.
- Въвеждане на обучителни програми за държавните служители относно антиромизма и онлайн речта на омразата.

За организациите на гражданското общество и образователните институции:

- Разработване на модели за гражданско дигитално участие, включително младежки мониторинг и кампании за повишаване на осведомеността.
- Осигуряване на емоционална подкрепа за млади хора, изложени на токсично съдържание.
- Подкрепа на инициативи, насърчаващи диалог, взаимно уважение и съвместни дейности между ромски и не-ромски младежи.
- Работа за промяна на нагласите чрез образование, медийна грамотност и изследване на механизма Стереотип - Предразсъдък - Дискриминация.

4.3 Европейски контекст

Първоначалните анализи показват, че онлайн антиромизмът не е ограничен до една държава, а следва общи европейски модели на дигитална дискриминация. Работата в рамките на проекта ТААО спомага за разработването на общи подходи, политики и инструменти за реакция. Окончателните изводи на европейско ниво ще бъдат добавени след завършването на всички национални доклади.

4.4 Ограничения

- Извадката не е представителна и отразява основно социалните мрежи и контактите на самите наблюдатели.
- Инструментите за докладване и реакциите на платформите варират, което усложнява сравненията между отчетните периоди.
- Могат да възникнат емоционални или интерпретативни пристрастия при декодиране на иронично или хумористично съдържание.

Въпреки тези ограничения, събраните данни предоставят ценна и систематизирана картина за формите, реториката и последиците от антиромската омраза онлайн.

4.5 Ключови изводи

Изследването подчертава необходимостта от решителен преход от реактивни към превантивни модели за противодействие на речта на омразата в дигиталната среда. Докато текущите подходи разчитат на послефактум докладване, премахване на съдържание или санкции, тези механизми адресират единствено видимите прояви на враждебност, без да се ангажират с по-дълбоките социални, когнитивни и културни двигатели на омразата. Ефективната превенция изисква цялостна рамка, в която технологичните инструменти, управлението на платформите и публичната политика се интегрират с дългосрочни образователни и общностни стратегии.

Техническите мерки - като усъвършенствани системи за модериране, алгоритмична прозрачност и филтри за ранно откриване - остават важни, но са недостатъчни сами по себе си. Техният ефект е ограничен, когато потребителите продължават да възпроизвеждат, споделят или пасивно толерират вредни наративи. Поради това превантивните усилия трябва да се допълнят с всеобхватни образователни инициативи, които засилват критичното мислене, насърчават етичното дигитално участие и повишават осведомеността за това как онлайн екосистемите формират възприятия и поведение.

Централен стълб на този превантивен подход е развиването на дигитална емпатия - способността да се разпознават, разбират и реагират на преживяванията на другите в онлайн среда. Дигиталната емпатия противодейства на деперсонализацията и емоционалната дистанцираност, които често позволяват на речта на омразата да се разпространява безпрепятствено. Тя подпомага конструктивния диалог, намалява импулсивните реакции и създава среда, в която дискриминацията и дегуманизацията са по-трудно нормализирани.

Не по-малко важно е насърчаването на самосъзнание относно ролята на личните стереотипи и имплицитни пристрастия. Много форми на онлайн враждебност не произтичат от явно намерение, а от непроверени предположения и усвоени социални модели. Разпознаването на начина, по който тези пристрастия влияят върху комуникационните практики, е от съществено значение за създаването на устойчиви промени в поведението. Когато индивидите разбират механизмите, чрез които стереотипите формират нагласите, те са по-добре подготвени да предизвикват вредни наративи,

да се противопоставят на манипулации и да участват отговорно в дигиталното пространство.

Всички тези елементи показват, че борбата с онлайн речта на омразата не е просто техническо или регулаторно предизвикателство, а по-широк обществен процес. Превенцията трябва да бъде закотвена в образованието, емпатията и колективната отговорност - условия, които дават възможност на индивидите и общностите да изградят по-инклузивни, устойчиви и безопасни дигитални среди.

Литература

Alliance against Antigypsyism. (n.d.). *Definition and framework of antigypsyism*. https://www.antigypsyism.eu

Barna, I., & Simonovits, B. (2020). *Hate speech and online radicalization: The role of algorithms in amplifying prejudice.* Journal of Digital Society, *5*(2), 21–38.

Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in America*. Rowman & Littlefield.

Breazu, P., & Machin, D. (2022). *Discourses of racism and nationalism in online media: A multimodal approach*. Routledge.

Council of Europe. (2022). *ECRI General Policy Recommendation No. 15 on combating hate speech*. Council of Europe.

European Commission. (2020). *EU Roma strategic framework for equality, inclusion and participation 2020–2030*. European Commission.

European Commission. (2021). *A more inclusive and protective Europe: Extending the list of EU crimes to hate speech and hate crime* (COM(2021) 777 final).

European Parliament. (2017). *Resolution on fundamental rights aspects in Roma integration in the EU: Fighting anti-Gypsyism* (2017/2038(INI)).

Goffman, E. (1986). *Frame analysis: An essay on the organization of experience*. Harper & Row.

Lentin, A. (2016). *Racism in public or public racism: Doing anti-racism in post-racial times. Ethnic and Racial Studies,* 39(1), 33–48. https://doi.org/10.1080/01419870.2016.1096409

Molnár, E. (2022). *Post-positivist grounded theory in qualitative media research*. European Journal of Qualitative Inquiry, *13*(2), 87–102.

PECAO – Peer Education to Counter Antigypsyist Online Hate Speech. (2022). *Final report*. European Roma Grassroots Organisation (ERGO) Network.

Rotaru, I., & Grebeldinger, D. (2025). *Together Against Antigypsyism Online (TAAO): Methodological framework and comparative synthesis report*. Asociatia Nevo Parudimos.

Siapera, E., Moreo, E., & Zhou, J. (2018). *HateTrack: Tracking and monitoring racist hate speech online*. Irish Human Rights and Equality Commission.

Smith, A. (2018). *Forms of counter-speech in digital contexts: Ethical considerations and practical approaches.* Journal of Digital Ethics, 7(3), 45–61.

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*. EU Kids Online. https://doi.org/10.21953/lse.47fdeqj01ofo

Commission for Protection from Discrimination. (2022). *Annual report on the activities of the Commission for Protection from Discrimination for 2021*. CPD. Retrieved from https://www.kzd-nondiscrimination.com

Ombudsman of the Republic of Bulgaria. (2022). *Annual report on human rights and non-discrimination in Bulgaria – 2021*. Ombudsman. Retrieved from https://www.ombudsman.bg

Center for the Study of Democracy. (2021). *Online propaganda, disinformation and hate speech in Bulgaria.* CSD. Retrieved from https://csd.bg

Bulgarian Helsinki Committee. (2023). *Human rights in Bulgaria in 2022: Annual report*. BHC. Retrieved from https://www.bghelsinki.org

Ministry of Interior. (2023). *Report on hate crimes in Bulgaria for 2022.* Ministry of Interior. Retrieved from https://www.mvr.bg

UNICEF Bulgaria. (2022). *Digital risks, online safety and the experiences of children in Bulgaria: National report 2022.* UNICEF. Retrieved from https://www.unicef.org/bulgaria

Open Society Institute – Sofia. (2021). *Public attitudes toward hate speech and discrimination in Bulgaria.* OSI–Sofia. Retrieved from https://osis.bg

Приложение 1: Общ инструмент за мониторинг

ИНСТРУМЕНТ ЗА МОНИТОРИНГ И АНАЛИЗ НА ПУБЛИКАЦИИ, СЪДЪРЖАЩИ РЕЧ НА ОМРАЗАТА В ИНТЕРНЕТ

I. ОБЩА ИНФОРМАЦИЯ

Социална медия	Тип акаунт				
	Лични/ частни акаунти	Онлайн медийни издания (новини, списания и др.)*	Институ-ционални публични акаунти**	Публични личности***	Инфлуенсъри в социалните медии
Facebook (Meta)					
Twitter (X)					
Instagram					
YouTube					
TikTok					
Друго (моля, уточнете)					

** Официална страница*

*** Правителствена или институционална организация (полиция, Министерство на образованието, държавни агенции и др.), общини, компании, неправителствени организации (включително църкви, училища, културни институции)*

**** Политици, известни професори/учени, артисти, музиканти, журналисти*

Име (на сайта/акаунта/изданието/канала)	
Дата на публикуване:	
Връзка или скрийншот	
Частен или публичен	
Автор (ако има такъв)	

Обхват (брой потребители/достигнати лица):

Харесвания	**Нехаресвания (включително „гняв“ или „тъга“)**	**Споделяния**	**Коментари**	**Преглеждания (Views / Визуализации)**

II. ТЕМАТИКА

ОСНОВНА ТЕМА	**ОБЯСНЕНИЕ**
Престъпления, извършени от роми	
Социални аспекти (жилищни условия, социални помощи, бедност, имиграция и др.)	
Образователни аспекти (отпадане от училище, условия на обучение, стипендии и др.)	
Здравни/санитарни аспекти (пандемия, достъп до болници, аборт и др.)	
Социални движения (протести, граждански права, представителство) и НПО	
Политика (представителство, политически партии, избори)	
Ромски лидери (включително жени)	
Културни събития (музика, филми, театър и др.)	
Спортни събития (състезания, игри)	
Друго (относно роми)	

По тристепенна скала, как оценявате общата рамка на темата:

Позитивно +	Негативно -	Неутрално 0

**Моля, умножете редовете толкова пъти, колкото е необходимо за всяка от идентифицираните теми.*

Тип съдържание	Позитивно +	Негативно -	Неутрално 0
Медийна информация (репортажи, новини и др.)			
Покана (културни/спортни събития, концерти, уебинари и др.)			
Обявления (информация за ежедневни дейности, прессъобщения)			
Гледна точка (редакционен тип, списания, впечатления, коментари, лични мнения)			
Реклама и препоръки (продажби, промоции, работни позиции, пътувания и др.)			
Любопитни факти (специални/ необичайни събития, истории)			
Забавление (музика, видеа, филми)			
Други			

Стил на съдържанието	Позитивно +	Негативно -	Неутрално 0
Емоционално			
Формално/Официално			
Призив за действие			
Забавно			
Артистично/фикционално/ фантастично			
Научно			
Други			

Съдържа ли визуални елементи?	
Да	**Не**

Моля, посочете използваните нетекстови форми (и техния брой, ако са повече от една)	
Снимки	
Миймове/ GIF-ове	
Карикатури/рисунки	
Мултимедийни материали – reels, story	
Видео	
Анимация	
Други	

По тристепенна скала, как оценявате общата рамка на визуалните елементи:

Позитивно +	**Негативно -**	**Неутрално 0**

**Моля, умножете редовете толкова пъти, колкото е необходимо, ако в публикацията има повече от една визуална форма.*

III. ОНЛАЙН ОМРАЗНА РЕЧ – ИНТЕНЗИВНОСТ И НИВА НА ОМРАЗАТА

Форми на явна онлайн омразна реч	Ниво/степен на омразата		
	1 приемлива	2 силна	3 екстремно силна
Груби епитети			
Расистки обиди			
Дехуманизиращи изказвания			
Персонални атаки (ad hominem)			
Наситени емоции (гняв, възмущение, враждебност)			
Подигравки и сарказъм			
Призив към насилие (включително убийство)			
Други			

Форми на закодирана онлайн омразна реч	Ниво/степен на омраза		
	1 приемлива	2 силна	3 екстремно силна
Използване на жаргон (ромски език)			
Ирония			
Псевдонаучни препратки (генетика, фалшиви статистики)			
Расово превъзходство – бял/черен			
Специфична реторика (метонимии, заобиколни изрази, двусмислия)			
Други			

Ориентир:

- ***Използване на жаргон*** *– използване на ромски език/думи за подчертаване на принадлежността към общността.*
- ***Ирония*** *– виж също сарказъм и подигравка („те не могат да отидат в рая, твърде тежки са, за да летят“ … заради колко бижута са откраднали).*
- ***Псевдонаучни препратки / фалшиви статистики*** *– използване на статистически данни, които не са официални, някои от тях идват от съмнителни научни сайтове или изследвания; например „80% от ромите не искат да работят“.*
- ***Whataboutery*** *(„а какво да кажем за нашите“) – препратка към „нас“ и „тях“, диалектичният общ възглас между „ние“ и „те“.*
- ***Метонимии*** *– изрази като „религия на измамата“, използвани вместо директното назоваване на ромите (без да се споменава думата точно).*
- ***Заобиколни изрази (circumlocutions)*** *– говорене около темата, например: „много мизерия в нашия район“, „толкова тъмно в този блок“.*
- ***Двусмислия (ambiguity)*** *– използване на пунктуация или формулировки, за да се направят реторични забележки, напр.: „трябва ли да бъдат изпратени в концентрационни лагери или не?!“.*

Съдържанието предава ли следните послания? До каква степен?

Статична срещу Динамична онлайн реч на омразата (Letin 2016)	Степени на токсичност на омразната реч			
	Пълно несъгласие	Частично несъгласие	Частично съгласие	Пълно съгласие
Затвърждаване на разделението/ поляризацията				
Подсилване и затвърждаване на стереотипите				
Разпространяване на митове/заблуди/ неистини				
Разпространяване на дезинформация/ фалшиви новини				
Оправдаване на изключването, стигматизацията и принизяването				
Затвърждаване на ексклузивистки национални идеи				
Други				

Ориентир:

- ***Укрепване на поляризацията*** = *„ние“ срещу „те“*
- ***Затвърждаване на стереотипите*** = *особено негативните (мързеливи, мръсни, неграмотни и др.)*
- ***Разпространяване на митове*** = *свръхестествени сили, магии и др.*
- ***Фалшиви новини*** = *например „те разпространяват болести чрез пътуванията си и начина си на живот“*

IV. ВИДОВЕ КОНТРАИЗКАЗВАНИЯ

1. Нарушава ли публикацията общоприетите норми на платформата? Да/Не
2. Какъв тип действия бихте предприели?

Начини за противодействие на омразната реч	Определено да	Не съм сигурен	Определено не
Искрено ангажиране			
Игнориране			
Разкриване на тактиката/стратегията на автора			
Оспорване на идеологията, вярванията и ценностите на автора			
Критикуване на аргументите и поведението на автора			
Подиграване или присмиване на съдържанието/на автора			
Отговаряне по същият начин			
Докладване на автора/съдържанието в системата на платформата			
Търсене на правна помощ от страна на гражданското общество			
Докладване на органите/ институциите			
Друго			

Ако решите да предприемете контрамерки, моля, предоставете доказателства за вашата дейност:

Вид доказателство	Дейност - кратко описание	Пример – скрийншот или линк
Текст – коментиране на поста		
Текст – докладване до платформата		
Текст – докладване до институция		
Визуално - скрийншот		

В случай, че има (само) коментари под публикацията/статията/изображението, предоставете някои подробности:

Анализ на коментарите	Честота	Примери (ако има такива)
Брой коментари		
Брой автори		
Тип на езиково изказване (агресивно vs. мирно) в 5степенна скала	Агресивно = 1 Мирно = 5	
Фалшив профил – скрита самоличност		
Интензитет на типа на коментарите	Липса на омраза = 0 Подигравка = 1 Омраза = 2 Подтикване на омразни действия = 3	
Устойчивост: настойчиво коментиране на публикацията с един или повече потребители	Неустойчиво = 0 Слабо = 1 Устойчиво = 2 Интензивно-устойчиво = 3	
Популярност и влияние на негативните коментари в тристепенна скала	Слаба = 1 Силна = 2 Екстремно силна = 3	

V. ДОКЛАДВАНЕ (ЛИЧНИ НАБЛЮДЕНИЯ):

Когато попълвате полето със собствено наблюдение, моля, опитайте се да се позовете на следните подточки (5–6 изречения, включително примери):

- **Негативни констатации за реалния проблем** – обяснение на проблемите/конфликта (обяснени ли са причините; посочено ли е кой е отговорен; има ли тенденция цялата вина да се прехвърля върху ромите);
- **Чии гледни точки са представени** – дали са показани и други гледни точки;
- **Какви стъпки и кога са предприети**;
- **Вашите препоръки** – към кого са насочени (платформата, авторът, институцията);
- **Какъв и кога е получен отговор** (посочете, ако не сте получили отговор);
- **Как е взето решението** – напр. бяха формулирани извинения, публикацията беше изтрита, авторът беше блокиран, авторът беше държан отговорен и др.;
- **Доволни ли сте от отговора/действията** – обяснете защо.

Попълнете полето с личните ви наблюдения за докладвания случай

Основна информация за докладвалия

Страна:	
Име на вашата организация:	
Данните са събрани от:	
Електронна поща:	
Дата на извършване на анализа:	
Дата на подаване на доклада:	

„Този доклад е изготвен в рамките на проект Together Against Antigypsyism Online (TAAO), финансиран от Европейския съюз чрез Европейската изпълнителна агенция за образование и култура (EACEA).

Funded by the European Union. Views and opinions expressed are however those ofthe author(s) only and do not necessarily reflect those of the European Union or European Education and Culture Executive Agency (EACEA). Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by
the European Union

TAAO

TOGETHER AGAINST ANTIGYPSYISM ONLINE