

Березень 2025

КЕЙСИ ТА ДОСЛІДЖЕННЯ

ІСТОРИЙ ВІЙНИ В УКРАЇНІ

Додаток до Посібника з управління ризиками в контексті надзвичайних ситуацій, збройних конфліктів та криз, заснованого на контекстному аналізі ситуації в Україні

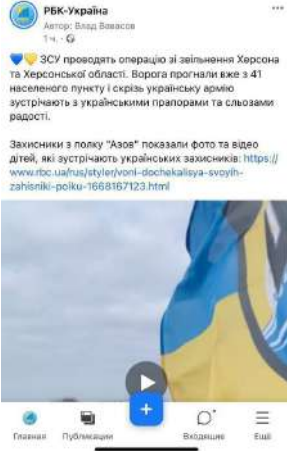


Кейси й дані дослідження

Наведені нижче кейси й дані дослідження репрезентують ризики, визначені в Посібнику з управління ризиками в контексті надзвичайних ситуацій, збройних конфліктів і криз (далі — Посібник). До них додані приклади реагування компаній-платформ соціальних медіа (далі — компанії) на виклики, пов'язані з повномасштабним вторгненням в Україну.


Кейси й дані дослідження містять рекомендації щодо конкретних ризиків. Це доповнення до тих рекомендацій, що наведені в Посібнику й пов'язані з реагуванням компаній на мову ворожнечі та дезінформацію, включно з тими, що спрямовані та впливають на жінок, меншини й маргіналізовані групи, у контексті майбутніх кризи, надзвичайної ситуації та збройного конфлікту. Також висвітлені ефективні дії, що їх виконали компанії, щоб адаптуватися до контексту надзвичайної ситуації, кризи та збройного конфлікту.

Кейс 1: Блокування та/або зменшення охоплення пов'язаного з війною контенту, який не порушує стандартів спільноти

Дата / період	Листопад 2022 р.
Платформа	Facebook
Опис	1 березня 2022 р. російські війська окупували українське місто Херсон. Після 256 днів окупації 11 листопада 2022 р. Збройні Сили України звільнили його. У листопаді на Facebook-сторінці інформаційного агентства «РБК-Україна» опублікували відео, де діти зустрічають українських бійців полку «Азов» у звільненому Херсоні. Після повідомлень користувачів, імовірно російських, механізми модерації Facebook згодом обмежили видимість цього та іншого контенту на сторінці РБК-Україна.
Реакція платформи	Facebook обмежив охоплення контенту зі сторінки РБК-Україна, посилаючись на порушення стандартів спільноти. Агентство звернулося до Facebook із проханням скасувати це рішення та попросило своїх читачів надіслати такі самі звернення. 19 січня 2023 р. міністр цифрової трансформації України Михайло Федоров повідомив: Meta заявила , що не блокуватиме контент про полк «Азов».
Джерело / додаткова інформація	У статті , опублікованій 22 листопада, РБК-Україна заявило, що стикнулося з обмеженнями контенту у Facebook через скаргу на пост зі згадкою про полк «Азов». На наведеному скриншоті — допис у Facebook РБК-Україна з відео, де херсонці вітають українську армію на звільненій території. Тут можна переглянути його онлайн-версію. 
Релевантність	У ситуаціях на кшталт збройного конфлікту в Україні, коли агресор застосовує тактику інформаційної війни, обмеження пов'язаного з війною контенту, який не порушує стандартів спільноти, може суперечити фундаментальному праву людини на доступ до інформації. Право на свободу вираження поглядів передбачає можливість громадськості шукати, отримувати та поширювати будь-яку інформацію, особливо під час кризи. Коли видаляють або обмежують контент, який дає критичне розуміння реалій конфлікту, такі дії не тільки обмежують множинність і різноманітність, але й

	<p>можуть ненавмисно посилити дезінформацію та мову ворожнечі, що зрештою нівелює зусилля щодо захисту права на правду та прийняття поінформованих рішень. Така модерація контенту може також сприяти негативним настроям, посилюючи невинуваті страхи та розгубленість серед цивільного населення і водночас погіршуючи нестабільну ситуацію. Крім того, перешкодження доступу до перевіреної інформації може призводити до втрати довіри до платформ, потенційно порушуючи право громадськості брати участь в інформованому публічному дискурсі. У цьому контексті модерація контенту має бути ретельно збалансованою, щоб запобігти подальшому маніпулюванню інформацією, водночас забезпечуючи захист прав людини та демократичних принципів.</p>
Рекомендації з урахуванням ризиків	<p>a) Збільшити зусилля щодо про підвищення обізнаності та навчання приватних користувачів, представників медіа, громадянського суспільства та інших відповідних стейкхолдерів щодо стандартів спільноти та політики модерації.</p> <p>b) Проводити постійний моніторинг як для алгоритмічного, так і для ручного прийняття рішень щодо можливого надмірно чи недостатньо суворого дотримання політики модерації. Регулярні оцінювання за участі місцевих стейкхолдерів допоможуть виявити проблеми.</p> <p>в) Підвищити прозорість під час розроблення та застосування політик модерації контенту, надаючи користувачам доступ до чітких і детальних пояснень щодо видалення чи зменшення охоплення їхніх публікацій. Наприклад, замість того, щоб обмежуватися загальною фразою на кшталт «це порушує стандарти спільноти», слід конкретно вказати, яку частину політики порушили і як контент суперечить відповідним нормам.</p> <p>г) Додати гендерно зумовлене насильство чи подібні речі до категорій повідомлень для користувачів.</p> <p>г) Відстежити, щоб механізми оскарження та перегляду, представлені українською мовою, були максимально доступними, зрозумілими та прозорими для користувачів.</p>

Кейс 2: Видалення контенту, який документує воєнні злочини

Дата / період	Червень 2023 р.
Платформа	Facebook, Instagram, YouTube
Опис	<p>У червні 2023 р. журналіст Ігор Захаренко, що документував наслідки російських нападів на мирних жителів у передмісті Києва, помітив, що його відео видаляють з акаунтів у Facebook, Instagram та YouTube. Матеріали, що містили докази воєнних злочинів, були автоматично видалені незабаром після їхнього завантаження. Так відбувалося через роботу алгоритмів, що їх соціальні мережі використовують для фільтрації контенту зі сценами насильства. Хоча ці алгоритми призначені для захисту користувачів від шкідливого контенту, вони часто не враховують контекст створення того чи іншого матеріалу, що призводить до видалення важливих доказів порушень прав людини.</p>
Реакція платформи	<p>Meta та YouTube заявили, що контент із зон бойових дій може залишатися на платформах, якщо це відповідатиме суспільним інтересам. Однак на практиці алгоритми часто видаляють такі матеріали автоматично. У Meta зазначили, що вони відповідають на запити правоохоронних органів у всьому світі та продовжують вивчати додаткові можливості для підтримки міжнародних процесів притягнення до відповідальності відповідно до юридичних зобов'язань і зобов'язань щодо конфіденційності.</p> <p>YouTube підкреслив, що попри існування винятків для контенту, який становить суспільний інтерес, платформа не є архівом, і рекомендував дослідникам, правозахисникам та журналістам використовувати інші способи збереження важливих матеріалів.</p>
Джерело / додаткова інформація	<p>У статті, опублікованій у червні 2023 р., BBC Україна навела приклади обґрунтувань, чому соціальні мережі видаляють докази воєнних злочинів в Україні. Матеріал збережений у форматі відеорепортажу на YouTube.</p> <p>На наведеному скриншоті зафіксоване подібне повідомлення з іншого кейсу, коли Instagram-допис українського журналіста, де були задокументовані воєнні злочини, видалили, оскільки механізми модерації платформи вважали його таким, що містить мову ворожнечі.</p> 

Релевантність	У контексті війни, що триває в Україні, документування воєнних злочинів, скоєних російськими та білоруськими військами, є фундаментальним питанням порушення прав людини. Дотримання права на справедливість, правду, репарації та неповторення потребує розслідування таких звірств і притягнення винних до відповідальності. Доступ до точного та своєчасного документування цих злочинів має вирішальне значення не лише для жертв, але й для української влади й міжнародної спільноти, які прагнуть забезпечити суверенітет України та захистити її від вторгнення. Хоча платформи мають свої процедури спілкування з правоохоронними органами, але вони не завжди ефективні, тож зрештою органи влади змушені звертатись до груп громадянського суспільства по доступ до інформації та видаленого контенту.
Рекомендації з урахуванням ризиків	<ul style="list-style-type: none"> а) Ініціювати співпрацю з місцевими й міжнародними організаціями, щоб забезпечити збереження видаленого контенту, який документує воєнні злочини, упродовж трьох — п'яти років. б) Зберігати опублікований користувачами контент із документуванням воєнних злочинів, який видається з платформ компанії, за межами України, у контексті з низьким рівнем ризику відповідно до міжнародних стандартів конфіденційності та захисту даних. в) Співпрацювати з правоохоронними органами та організаціями громадянського суспільства, які мають досвід архівування та анонімізації даних, щоб покращити й оприлюднити параметри ідентифікації критично важливого контенту щодо прав людини, який має бути збережений, а також процеси сортування контенту для надання правоохоронним органам. г) Інформувати користувачів, чи зберігатимуть компанії їхній контент. г) Поінформувати відповідних місцевих стейкхолдерів про існування архівних матеріалів, які стосуються процесів притягнення до відповідальності та правосуддя за воєнні злочини. д) Приділяти особливу увагу злочинам проти жінок, меншин і маргіналізованих груп, пов'язаному з конфліктом сексуальному насильству та співпраці з експертними правозахисними організаціями.

Кейс 3: Використання ботів і фейкових акаунтів для поширення дезінформації та мови ворожнечі


Дата / період	Листопад 2023 р.
Платформа	TikTok
Опис	У листопаді 2023 р. у численних акаунтах TikTok з'явилися відео, у яких обіцяли вигадані грошові виплати від 9 000 до 50 000 грн (приблизно 200–1 200 доларів США) для жителів, які залишилися в Україні після початку повномасштабного вторгнення. Користувачам пропонували переходити за посиланнями, де вони підписувались на канали анонімних ботів у Telegram. Ці канали містили дезінформацію, фіктивні опитування на політичні теми й неперевірену інформацію про війну в Україні, наприклад підроблені урядові рішення.
Реакція платформи	На момент проведення дослідження відповідь на запит щодо ідентифікованих акаунтів від TikTok не надходила.
Джерело / додаткова інформація	<p>Інститут масової інформації (ІМІ) представив дослідження поширення дезінформації українською мовою в TikTok у статті, опублікованій у січні 2024 р. У ній задокументували, як фейкові канали TikTok використовують, щоб скеровувати користувачів на анонімні Telegram-канали.</p> <p>Більшість виявлених фейкових відео в TikTok складаються з уривків новинних програм популярних українських телеканалів. На наведеному скріншоті вказана неправдива інформація про міжнародну допомогу Україні (7 000 грн від США), про яку нібито розповідають на популярному телеканалі 1+1. Тут можна переглянути його онлайн-версію.</p>
Релевантність	<p>Експлуатування страхів і невпевненості українців, що в цьому кейсі пов'язані з фінансовими труднощами, за допомогою дезінформації та шахрайства є порушенням права громадськості на точну та правдиву інформацію. Поширення шкідливого контенту не лише впливає на здатність людей приймати поінформовані рішення, але й ставить під загрозу психічний добробут і безпеку українців, зокрема представників вразливих категорій, як-от молодь та соціально-економічно незахищені групи. Цілеспрямована дезінформація може руйнувати довіру до життєво важливих державних інституцій та органів влади, водночас послаблюючи соціальну єдність і порушуючи право жити в мирному та безпечному середовищі.</p>



	<p>Та дезінформація, що грає на страхах людей, використовує їхню вразливість і посилює почуття тривоги, паніки та розгубленості, що може призвести до зростання суспільної напруги. У цьому контексті відповідальність за захист людей від шкідливого контенту лежить як на державних органах, так і на приватних платформах. Такі платформи, як TikTok, повинні забезпечити для користувачів ефективний захист прав людини, запобігаючи поширенню таких шкідливих матеріалів, особливо серед тих, хто найбільше ризикує бути введеними в оману. Водночас уряди повинні забезпечити, щоб політика та дії захищали право на інформацію, сприяли медіаграмотності та створювали стійкість серед вразливих спільнот для боротьби з дезінформацією та запобігання її негативному впливу на суспільство.</p> <p>Хоча боти у цьому кейсі були локалізовані в Telegram, важливі виклики щодо штучної діяльності та ботів, а також вразливості під час їхнього виявлення помітили на всіх основних платформах соціальних медіа.</p> <p>Задokumentовані кейси ботів і тролів, які коментують і лайкають дописи, що підтримують дезінформаційні наративи, або атакують перевірений контент і дані про російську агресію, і в такий спосіб створюють ілюзію масової підтримки певних поглядів. Дезінформація зазвичай вибудовується на шкідливих стереотипах, які часто є виразно гендеризованими та сексуалізованими. Вона має на меті розпалювати суспільну поляризацію та суперечки й спрямована та впливає особливо на жінок, меншини та маргіналізовані групи. Також фейкові акаунти, які нібито належать українським військовослужбовцям чи волонтерам, поширюють маніпулятивну й неперевірену інформацію про втрати української армії, неправдиві новини про гуманітарні катастрофи. Такі повідомлення сіють паніку та відчай і підривають довіру до українських інституцій.</p> <p>Ця маніпуляція не тільки порушує право особи на безпеку, але й посягає на її гідність та безпеку від шкоди. Як уряди, так і платформи відповідають за захист людей і мають запобігати поширенню дезінформації, сприяти медіаграмотності та забезпечувати реалізацію права на перевірнену інформацію, особливо під час кризи.</p>
<p>Рекомендації з урахуванням ризиків</p>	<ul style="list-style-type: none"> a) Пріоритезувати ресурси для визначення, модерування й блокування підроблених і автоматизованих акаунтів, сторінок і груп, які публікують і поширюють дезінформацію та мову ворожнечі, особливо гендерну дезінформацію та дискримінаційний контент. б) Створити ефективні системи раннього попередження та канали ескалації за участі місцевих стейкхолдерів, наприклад надійних фактчекерів, медіа, експертів з гендерних питань та організацій громадянського суспільства, для надзвичайних ситуацій штучної масової діяльності, що сприяє особливо небезпечній дезінформації та нарративам ворожнечі, включно з тими, що розпалюють упередження й поляризацію суспільства і спрямовані на жінок, меншини та маргіналізовані групи. в) Проводити регулярний моніторинг та оцінювання алгоритмічного й ручного прийняття рішень у звітах компанії про прозорість. Основні висновки поширювати та обговорювати з місцевими стейкхолдерами. г) Налагодити співпрацю з національними та місцевими стейкхолдерами, які мають можливості та потенціал для продукування інформації, що пройшла фактчекінг, та протидії шкідливому й оманливому контенту. Надати цим стейкхолдерам потрібний доступ, аби вони мали змогу висувати пропозиції щодо подолання кризи й оперативно реагувати, щоб удосконалити набори даних, на яких побудовані алгоритми модерування. г) Контент, що пройшов фактчекінг і має на меті протидіяти дезінформаційним нарративам, ідентифікувати та просувати відразу після сплесків масової діяльності, що пропагує шкідливі наративи, і пропорційно до цієї діяльності. Особливу увагу приділяти дезінформації, яка порушує права людини, включно з правами жінок, і шкодить боротьбі з дискримінацією. д) Пов'язаний з війною контент, що пройшов фактчекінг і відповідає суспільному інтересу, наприклад пов'язаний з питаннями безпеки та гуманітарною допомогою, ідентифікувати та просувати у співпраці з кризовими експертами та місцевими стейкхолдерами. е) Приділяти особливу увагу реальним акаунтам з довгою історією, які можуть бути «здані в оренду» для поширення дезінформації. Виявляти такі випадки шляхом моніторингу раптового зростання підозрілої активності та відстеження нетипової поведінки акаунту.

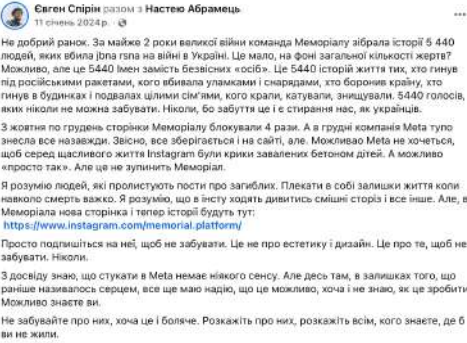
Кейс 4: Наявність неправдивого, оманливого та зловмисного контенту (зокрема згенерованого ШІ), та контенту, націленого на жінок, меншини та маргіналізовані групи

Дата / період	Травень 2023 р.
Платформа	Google
Опис	У травні 2023 р. виявили мережу з 49 новинних сайтів, які містили штучно створений контент. Сайти, які мали версії різними мовами, розміщували статті, що поширювали дезінформацію та пропагували війну Росії проти України. Один із них — TNewsNetwork — опублікував неправдиву історію про загибель тисяч солдатів на війні, засновану лише на відео в YouTube, що не містило підтверджених фактів.

Реакція платформи	Після проведеного розслідування Google видалив рекламу з певних сторінок цих сайтів. У випадку систематичних порушень рекламу блокували на всьому сайті. Однак наявність згенерованого ШІ контенту автоматично не є порушенням рекламної політики. Але якщо автоматизацію, включно з ШІ, використовують для маніпулювання рейтингом у результатах пошуку, то це вважається порушенням політики спаму.
Джерело / додаткова інформація	<p>У статтях, опублікованих навесні 2023 р. у Bloomberg та NewsGuard, показано, як інструменти ШІ використовують для створення контенту на так званих контент-фермах. Це сайти низької якості з великою кількістю згенерованого ШІ контенту, який не проходить фактчекінг і часто поширює неправдиву інформацію.</p> <p>На наведеному нижче скриншоті — згенерований ШІ заголовок, який з'явився на TNewsNetwork.com, анонімному новинному сайті, зареєстрованому в лютому 2023 р.</p> 
Релевантність	<p>Згадані в кейсі сайти часто відповідають формальним вимогам платформ соціальних медіа, тому можуть брати участь у рекламних програмах і отримувати прибуток, поширюючи дезінформацію.</p> <p>Автоматизація сприяє швидкому поширенню дезінформації, яку використовують для розпалювання недовіри, суспільної поляризації, поширення шкідливих стереотипів і гендерного насильства. Зокрема, це може бути контент, спрямований на жінок, меншини й маргіналізовані групи, що поширює неправдиві наративи про українських військових та інституції. Були задокументовані приклади дипфейкових відео із зображенням українських політиків, які створили, щоб спровокувати паніку та підірвати довіру до демократичних інституцій.</p> <p>Платформи соціальних медіа можуть посилювати гендерну дезінформацію та мову ворожнечі, особливо спрямовану проти жінок, які займають публічні посади, як-от політикині, журналістки, правозахисниці та ЛГБТК-активістки. Такі інформаційні кампанії мають на меті підірвати довіру до жінок-лідерів шляхом поширення неправдивої інформації, яка ставить під сумнів їхні компетентність, моральні якості чи лояльність до держави. Цей тип онлайн-насильства посилює гендерну нерівність, порушує право на свободу вираження поглядів і право осіб, особливо жінок, представників меншин і маргіналізованих груп, брати участь у суспільному житті без дискримінації та шкоди.</p>
Рекомендації з урахуванням ризиків	<ol style="list-style-type: none"> Запровадити механізми тимчасового обмеження або контрольованого скорочення охоплення контенту, який має ознаки неправдивої, оманливої або зловмисної інформації, зокрема інформації, створеної ШІ. Водночас особливу увагу приділяти гендерній дезінформації та дискримінаційному контенту. Створити ефективні системи раннього попередження й канали ескалації для надзвичайних ситуацій, які можуть вплинути на фізичну безпеку осіб. Пріоритизувати виділення достатніх фінансових і людських ресурсів для виявлення аудіо-, візуального та аудіовізуального контенту, створеного ШІ, і для розробки нових технологій, які підвищують точність такої ідентифікації. Запровадити маркування штучно створеного текстового, аудіо-, візуального та аудіовізуального контенту.

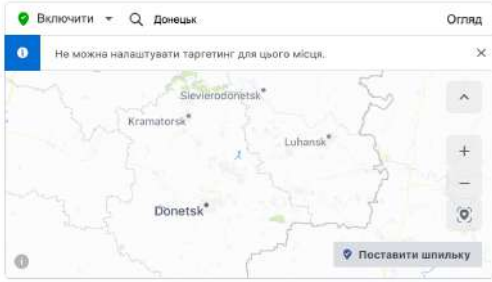
Кейс 5: Політики та практики модерації, що не враховують розуміння лінгвістичного, соціального, політичного, історичного й культурного контексту, зокрема щодо меншин, гендерних і маргіналізованих груп

Дата / період	Січень 2024 р.
Платформа	Instagram
Опис	У період із жовтня до грудня 2023 р. Instagram-сторінки Меморіалу — проекту, що збирає історії про полеглих та вбитих воїнів і цивільних осіб , блокували чотири рази, а в грудні 2023 р. компанія «Мета» видалила акаунти з Instagram. З березня 2022 р. на цих сторінках публікували історії про загиблих під час війни українських воїнів. Через видалення організатори проекту втратили доступ до 5 440 детальних акаунтів, створених для збереження пам'яті про воїнів.

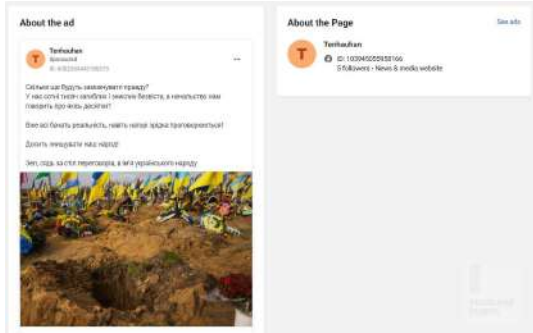
Реакція платформи	Після хвилі підтримки користувачів із закликком розблокувати акаунт Meta відновила Instagram-сторінки Меморіалу в січні 2024 р.
Джерело / додаткова інформація	<p>Незалежний ресурс «Детектор медіа» — експертна організація з протидії дезінформації та пропаганді — у статті, опублікованій у січні 2024 р., задокументував відновлення Instagram-акаунтів Меморіалу після протестів українських користувачів. Тут можна переглянути збережену версію статті.</p> <p>На наведеному нижче скриншоті — допис у Facebook координатора проєкту «Меморіал», у якому він просить спільноту підписатись на нову сторінку в Instagram.</p> 
Релевантність	<p>Блокування популярних акаунтів з перевіреною інформацією позбавляє українців права висловлювати скорботу та вшановувати загиблих воїнів і мирних жителів, обмежує обізнаність громадськості про жертви війни та послаблює підтримку постраждалих родин. Такі дії також підривають довіру до платформ соціальних медіа, особливо коли користувачі стикаються з видаленням контенту або призупиненням активності акаунту за критику пропаганди чи жарти, пов'язані з місцевою культурою. Це обмежує свободу вираження поглядів українців, заважає їм брати участь у важливих публічних дебатах і ділитися власним досвідом. Важлива проблема — це неспроможність платформ розпізнати делікатний контекст місцевих висловів, що створює ризик дискримінації жінок, меншин і маргіналізованих груп.</p> <p>Видалення або приховування контенту, який документує реалії війни, може перешкоджати праву на доступ до перевіреної інформації, яке є вкрай важливим для можливості громадськості приймати поінформовані рішення. Коли платформи віддають перевагу модерації контенту, а не збереженню критично важливої перевіреної інформації, вони невідомо сприяють циклу дезінформації, роблячи користувачів вразливими до фальшивих наративів. Під час війни така цензура може підрвати суспільну довіру, дестабілізувати соціальну згуртованість і заглушити голоси, які є життєво важливими для проживання горя, притягнення до відповідальності та прогресу.</p>
Рекомендації з урахуванням ризиків	<ol style="list-style-type: none"> Інтегрувати контекстуальний аналіз у модерацію на всіх рівнях, забезпечуючи співпрацю з лінгвістами та експертами з соціокультурних особливостей і регіональної специфіки. Аналіз має бути гендерно чутливим і робити акцент на контекстних дискримінаційних дискурсах та пов'язаних термінах. Удосконалити кваліфікацію та підготовку модераторів, зосередившись на поглибленні їхнього розуміння культурних, історичних, гендерних та політичних контекстів. Узгодити різні аспекти модерації й унеможливити прийняття рішень алгоритмами без реального аналізу контексту.

Кейс 6: Брак можливості охоплювати користувачів на тимчасово окупованих територіях України

Дата / період	2022 р.
Платформа	Facebook
Опис	Українські медіа зіткнулися з обмеженнями щодо контенту, орієнтованого на тимчасово окуповані території (ТОТ). Це унеможливило таргетування на Крим і Севастополь через правові обмеження, пов'язані з санкціями.
Реакція платформи	Meta пояснила, що обмеження необхідні для дотримання міжнародних санкцій, але не запропонувала альтернативних способів інформування жителів окупованих регіонів.



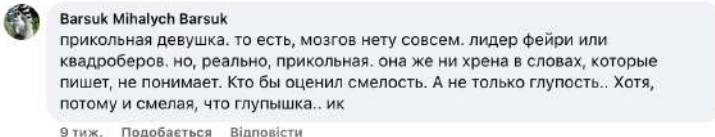
<p>Джерело / додаткова інформація</p>	<p>У статті, опублікованій у липні 2022 р., інтернет-газета «Українська правда» задокументувала, як рекламні обмеження компанії Meta на окупованих українських територіях перешкоджають здатності України протистояти російській пропаганді й надавати правдиву інформацію цивільному населенню.</p> <p>На наведеному нижче скриншоті показана неможливість таргетування на користувачів у Донецькій області.</p> 
<p>Релевантність</p>	<p>Мешканці ТОТ опинилися в інформаційному вакуумі, що підвищило їхню вразливість до російської дезінформації та ускладнило доступ до перевіреної інформації про події в Україні. За таких умов дезінформація може легко поширюватися, маніпулюючи громадською думкою та викликаючи розгубленість щодо ситуації на місцях.</p> <p>Співпраця з медіа, громадянським суспільством, органами влади чи іншими відповідними стейкхолдерами, щоб націлитися на користувачів у регіонах, вразливих до дезінформації та з обмеженим доступом до інформації, може мати критичне значення для безпеки місцевих мешканців. Охоплення користувачів у ТОТ є життєво важливою частиною захисту їхнього права на доступ до правдивої, перевіреної інформації, що дає їм змогу залишатися на зв'язку із зовнішнім світом і приймати поінформовані рішення. Такі зусилля також запобігають поширенню шкідливої пропаганди та допомагають мешканцям протистояти маніпуляціям, зміцнюючи їхню здатність захищати свої права в нестабільному середовищі.</p>
<p>Рекомендації з урахуванням ризиків</p>	<p>a) Налагодити співпрацю з місцевими стейкхолдерами, включно з представниками ТОТ, щоб обговорювати й шукати рішення для надання місцевим жителям на тимчасово окупованих територіях перевіреної інформації. Такі рішення мають враховувати потреби жінок, меншин, маргіналізованих і вразливих груп.</p>

Кейс 7: Зловживання комерційними інструментами в політичних та воєнних цілях, що порушують стандарти спільноти

<p>Дата / період</p>	<p>2022 р.</p>
<p>Платформа</p>	<p>Facebook</p>
<p>Опис</p>	<p>У 2022 р. у Facebook виявили організовані кампанії з фейковими сторінками, які використовували платну рекламу для поширення ідентичних текстів та різних візуалізацій, що містили наративи політичної дезінформації.</p>
<p>Реакція платформи</p>	<p>Facebook виявив і видалив мережі фейкових акаунтів та сторінок, які порушували політику платформи щодо автентичної поведінки.</p>
<p>Джерело / додаткова інформація</p>	<p>Незалежний розслідувальний ресурс «Реальна газета» у статті, опублікованій у березні 2023 р., детально описав, як Росія використовує рекламу у Facebook для поширення неправдивої інформації, що деморалізує, серед українських аудиторій через недовговічні анонімні сторінки. Тут можна переглянути збережену версію матеріалу.</p> <p>На наведеному скриншоті — один із прикладів таких сторінок. На сторінці стверджують, що українська влада приховує кількість загиблих на війні українців.</p> 

<p>Релевантність</p>	<p>Дезінформаційні кампанії, спрямовані на послаблення морального духу та дискредитацію влади, можуть вводити громадськість в оману, посилювати страхи та недовіру до офіційних джерел інформації, створювати непотрібне навантаження на органи влади та демократичні інститути. Це, зокрема, сприяє дестабілізації суспільства, що є особливо критичним під час надзвичайних ситуацій, криз та збройних конфліктів. За допомогою інструментів на кшталт платної реклами та цільового й автоматичного просування контенту можна значно підвищити ефективність дезінформаційних кампаній та охоплення шкідливого контенту.</p> <p>Спонсоровані дописи чи рекламні кампанії, наприклад просування наративів про масові втрати, мають на меті підірвати моральний дух і деморалізувати громадян. Такі кампанії не лише дестабілізують суспільство, але й загрожують інформаційній безпеці, поширюючи емоційний стрес і фейки. Ці дії порушують право громадян на правду, загрожують громадській безпеці та послаблюють здатність людей приймати поінформовані рішення в критичних ситуаціях.</p>
<p>Рекомендації з урахуванням ризиків</p>	<p>a) Збільшити зусилля щодо забезпечення прозорого оприлюднення інформації про рекламні кампанії та покупок реклами.</p> <p>b) Провести розслідування, щоб переконатися, що програми монетизації не спрямовують доходи акторам, пов'язаним із суб'єктами, що перебувають під санкціями, або іноземним акторам, які систематично створюють та/або поширюють дезінформацію й наративи ворожнечі.</p> <p>v) Розробити програму незалежного аудиту рекламних кампаній із залученням зовнішніх експертів та організацій громадянського суспільства.</p>

Кейс 8: Пов'язані з гендерними питаннями дезінформація та мова ворожнечі

<p>Дата / період</p>	<p>Грудень 2024 р.</p>
<p>Платформа</p>	<p>Facebook</p>
<p>Опис</p>	<p>19 грудня 2024 р. Україна зазнала масштабної кібератаки на державні реєстри, що призвело до їх тимчасової недоступності. Згодом у соцмережах, зокрема у Facebook, виникла хвиля негативних коментарів та мови ворожнечі проти співзасновниці Фондації жіночого лідерства та стратегічних ініціатив. Атаки посилювалися після її участі в заході, присвяченому ролі жінок у кібербезпеці, де вона публічно висловила свою позицію щодо посилення гендерного балансу в галузі технологій.</p> <p>Напади мали яскраво виражений гендерний характер. Коментарі зосереджувалися навколо ворожих і гендерно-стереотипних тверджень про брак компетенції за ознакою статі. Вульгарні та принизливі висловлювання, не пов'язані з її професійною діяльністю, були спрямовані на дискредитацію жінки як експерта.</p>
<p>Реакція платформи</p>	<p>Щоб розв'язати ситуацію, не вжили жодних ефективних заходів. Скарги жінки, про яку мова, та інших осіб не спричинили видалення контенту.</p>
<p>Джерело / додаткова інформація</p>	<p>Допис у Facebook у грудні 2024 р. спричинив гендерні напади та мову ворожнечі після того, як співзасновниця Фондації жіночого лідерства та стратегічних ініціатив взяла участь у заході, присвяченому ролі жінок у кібербезпеці. Нижче наведені приклади публікацій. Більше можна знайти в оригінальному дописі у Facebook.</p> <p> Roman Kurbatov 20 грудня 2024 р. ...</p> <p>Mike Sydorenko, посмотри какое вкусное. Если у нас вся кибербезопасность только этим и занята, я вообще не удивлён тому, что мы остались без реестров.</p> <p>«(...) подивись, яке смачне. Якщо у нас вся кібербезпека тільки цим і зайнята, я взагалі не вражений тим, що ми zostалися без реєстрів».</p> <p> Roman Budzyk Усі порти у мадам будуть славно від'юзані...без термопасти і охолодження ...</p> <p>9 тиж. Подобається Відповісти</p> <p>«Усі порти у мадам будуть славно від'юзані ... без термопасти і охолодження»</p> <p> Barsuk Mihalych Barsuk прикольная девушка. то есть, мозгов нету совсем. лидер фейри или квадроберов. но, реально, прикольная. она же ни хрена в словах, которые пишет, не понимает. Кто бы оценил смелость. А не только глупость.. Хотя, потому и смелая, что глупышка.. ик</p> <p>9 тиж. Подобається Відповісти</p> <p>«прикольна дівчина, тобто мізків немає зовсім. лідер фейрі чи квадроберів, але, реально, прикольна. вона ж ні чорта в словах, які пише, не розуміє. Хто б оцінив сміливість. А не тільки дурість... Хоча, тому й смілива, що дуренька... ік»</p>

Релевантність	Брак належних інструментів модерування та реагування призводить до посилення дискримінації та загрожує безпеці. Захист жінок, меншин і маргіналізованих груп від мови ворожнечі та дезінформації має важливе значення для дотримання прав людини, запобігання шкоді та насильству, а також сприяння згуртованості суспільства. Ці групи часто стають об'єктами непропорційного таргетування, і неконтрольована мова ворожнечі може перерости в реальну шкоду. Гарантування їхньої безпеки в інтернеті сприяє рівності, свободі вираження думок і здоровому цифровому середовищу, де всі учасники не відчувають страху. Це також допомагає підтримувати довіру громадськості, заохочує її участь і зміцнює цілісність цифрових платформ як простору для змістовного дискурсу.
Рекомендації з урахуванням ризиків	<ul style="list-style-type: none"> а) Співпрацювати з місцевими, регіональними та міжнародними експертами для інформування та вдосконалення політики та практики модерації, щоб захистити жінок, меншини й маргіналізовані групи. б) Забезпечити систему підтримки для жінок, меншин і маргіналізованих груп, які стикаються з гендерною дезінформацією та мовою ворожнечі, наприклад механізми швидкого інформування, доступ до ресурсів для зміцнення психічного здоров'я та варіанти захисту акаунтів. в) Співпрацювати з неурядовими організаціями та місцевими жіночими організаціями для підтримки й захисту жінок, що перебувають у центрі уваги громадськості, від потенційної фізичної шкоди, пов'язаної з онлайн-зловживаннями.

Дослідження

Є ризики, які не були проілюстровані конкретними випадками, але становлять загрозу для доступу до інформації й однаково релевантні з погляду їхнього розв'язання. Це два з десяти ключових ризиків, визначених місцевою групою експертів для Посібника з управління ризиками в контексті надзвичайних ситуацій, збройних конфліктів і криз. Нижче наведені дані незалежного дослідження, що дають змогу зрозуміти основні виклики.

Неефективні інструменти для пошуку на платформах перевіреної інформації, пов'язаної з війною

Незалежний український медіаресурс [«Texty.org.ua»](https://texty.org.ua) зафіксував, що під час війни українці мали неефективні інструменти для пошуку на платформах перевіреної інформації, пов'язаної з війною. Влітку 2023 року 205 українців зголосилися дозволити медіаресурсу моніторити рекомендації, які надавав YouTube.

Ресурс дійшов висновку, що існує кілька проблем, пов'язаних з тим, що YouTube рекомендує українським користувачам російську дезінформацію та мову ворожнечі. Однією з основних проблем є те, як алгоритм платформи просуває контент на основі мови. Так, користувачеві («Vatnik»), який переглянув рекомендоване відео проросійського блогера, згодом порадили більше відео та каналів, які пропагують дезінформацію про війну в Україні. Рекомендації містили інтерв'ю та ролики з американським псевдоекспертом з проросійськими поглядами Скоттом Ріттером, два YouTube-акаунти якого платформа закрила за поширення неправдивої інформації. Тобто YouTube просував осіб та організації, яких сам забороняв в інших випадках. Ще однією виявленою проблемою було те, що проросійську дезінформацію часто відносили до неправильної категорії, тож користувачам, які, наприклад, шукали музику, рекомендували політичний контент і дезінформацію про війну.

Усі ці виклики, пов'язані з рекомендаціями YouTube, ускладнюють для українців, які перебувають в умовах інформаційної війни, пошук і легке знаходження перевіреної інформації, пов'язаної з війною. Це суперечить їхньому праву на доступ до інформації у вразливий період.

Рекомендації щодо конкретного ризику:

- а) Продовжувати співпрацю з місцевими медіа та організаціями й установами, що спеціалізуються на підвищенні інформаційної грамотності, включно з організаціями, сфокусованими на гендерно чутливому підході та ознайомленими з місцевим гендерним контекстом.
- б) Запровадити маркування акаунтів, які мають ознаки штучної скоординованої поведінки. Воно може ґрунтуватись на кількох індикаторах, наприклад використання ботів для взаємодії з контентом, аномальні сплески активності, невідповідність географічного розташування, зміна назви, а також аналіз поведінки акаунту, включно зі «здачею в оренду» для маніпулятивних дій.
- в) Збільшити фінансові й людські ресурси для ефективного та кваліфікованого розгляду апеляцій від користувачів, яких помилково визначили як ненадійне джерело. Залучити місцевих стейкхолдерів до оцінювання ключових кейсів.

Рекомендації користувачам шкідливого контенту, включно з дезінформацією та мовою ворожнечі

Згідно з результатами дослідження, проведеного в листопаді 2024 року Центром стратегічних комунікацій та інформаційної безпеки, TikTok активно використовують для поширення дезінформації серед української аудиторії. Алгоритми платформи просувають дезінформацію з проросійськими наративами, наприклад контент, що дискредитує українську армію. Відео, що містять фейкову інформацію про мобілізацію, неправдиву критику української влади та заперечення воєнних злочинів, отримують велику кількість переглядів завдяки механізмам автоматичних рекомендацій.

Прикладом є антиукраїнський та проросійський інфлюенсер у соціальних мережах Антон Гура, який має 48 200 підписників. Задокументовано, що він поширював дезінформацію про впровадження електронної гривні та відключення електроенергії в Україні, які нібито були спричинені конфліктом між урядом та національними інфраструктурними компаніями, а не російськими атаками. Його найпопулярніші відео на TikTok — це відео, спрямовані проти мобілізації. Вони набирають від десятків тисяч до 1,5 мільйона переглядів.

Відео Гури є частиною масштабнішого тренду. З 16 травня до 16 червня 2024 року опублікували орієнтовно 13 000 відео з хештегом #ТСС, які загалом набрали понад 470 мільйонів переглядів. ТСС розшифровується як Територіальний центр комплектування та соціальної підтримки. На ньому сфокусовані численні російські дезінформаційні кампанії, що мають на меті зірвати мобілізацію до української армії. У липні #ТСС став найпопулярнішим хештегом в українському сегменті TikTok разом з іншими антимобілізаційними хештегами, як-от #stoptsk, #poprattik, #spolotivtsek, #protesttsk, а також з антидержавним #senomeyUkraine.

Іншим прикладом є дописи, у яких президента Росії Володимира Путіна — особу, щодо якої після розслідування воєнних злочинів, злочинів проти людяності та геноциду Міжнародний кримінальний суд (МКС) видав ордер на арешт, — вихваляють як найвеличнішого лідера світу. Їх рекомендують у стрічках українців. Такий контент може слугувати виправданням злочинних дій. Це знецінює важливість притягнення до відповідальності та пошук правди для жертв порушень прав людини й не захищає гідність українців, які потерпіли від російського вторгнення.

Рекомендації щодо ризику:

- а) Співпрацювати з місцевими медіа, групами громадянського суспільства та іншими відповідними стейкхолдерами, щоб підтримувати розвиток медіаграмотності серед цільових аудиторій.
- б) Створювати спільноти з місцевими експертами, які здатні завчасно попереджати про шкідливий і рекламний контент та реагувати на його сплески, а також пропонувати рекомендації щодо модерації.
- в) Проводити тренінги для модераторів компаній, де звертати увагу на виявлення гендерно мотивованої дезінформації та гендерно зумовленого насильства.

Проект втілюється організацією IMS (International Media Support) та ГО «Інтерьюз-Україна» у партнерстві з UNESCO та за підтримки Японії. Проект ґрунтується на «Посібнику з регулювання цифрових платформ» від 2023 року.
