Березень 2025

ПОСІБНИК

з управління ризиками в контексті надзвичайних ситуацій, збройних конфліктів та криз

На основі контекстного аналізу ситуації в Україні

unesco

From the People of Japan

IMS

ІНТЕРНЬЮЗ УКРАЇНА

Узагальнені рекомендації

1. **Адаптація стандартів спільноти та принципів модерації**

   Розглянути можливість гнучкого застосування своїх глобальних правил та стандартів, а також практик модерації в регіонах, що перебувають в умовах надзвичайних ситуацій, криз та збройного конфлікту, орієнтуючись на норми міжнародного гуманітарного права, права людини та місцевий досвід. Тобто використовувати уніфіковані підходи для кластерів країн зі спільними ризиками, зокрема пов'язаними з реалізацією права на самооборону та самовизначення. Така гнучкість має передбачати залучення відповідних місцевих зацікавлених сторін (стейкхолдерів) та відстеження динамічних подій у конфліктних регіонах на кшталт України, щоб платформи могли своєчасно оновлювати налаштування алгоритмів і уникати невиправданого блокування або зменшення видимості контенту.

2. **Створення регіональних кризових команд**

   За можливості додати до кризових протоколів пункт про формування спеціалізованих команд для оперативного реагування на події, пов'язані з кризами, надзвичайними ситуаціями та збройним конфліктом. У співпраці з місцевими організаціями такі команди можуть забезпечити швидке прийняття рішень та адаптацію політик і стандартів спільноти до контексту. На прикладі України активісти продемонстрували готовність брати активну участь у боротьбі з дезінформацією та мовою ворожнечі.

3. **Регулювання використання комерційних інструментів**

   Підвищити прозорість використання рекламних інструментів, щоб запобігти маніпуляціям і зловживанням. Підвищення прозорості фінансування реклами, аналізування її змісту та джерел і зрештою зниження ризиків поширення дезінформації сприятимуть зміцненню довіри до платформ соціальних медіа.

4. **Підтримання та зміцнення зусиль щодо перевірки фактів**

   Докладати більше зусиль щодо перевірки фактів (фактчекінгу) та співпраці. Незалежні дослідники[1] та експерти[2] дійшли висновку, що фактчекінг має важливе значення для протидії поширенню дезінформації. Варто не лише зберегти наявні механізми фактчекінгу, але й активно розвивати та адаптувати їх до місцевих потреб, забезпечуючи співпрацю з локальними фактчекерами й експертами з урахуванням регіональної специфіки.

[1] Van Erkel, P. F. A. et al., (2024).
[2] European Commission, (2018).

Посібник з управління ризиками має на меті надати вказівки та рекомендації щодо пом'якшення наслідків надзвичайних ситуацій, збройних конфліктів та криз для компаній-платформ соціальних медіа (далі — компанії). Посібник пропонує контекстні межі, які дають змогу компаніям захищати права людини, зокрема свободу слова та доступ до інформації, в українській інформаційній екосистемі. Водночас матеріал акцентує на таких категоріях, як-от жінки, меншини й маргіналізовані групи, які зазвичай є таргетованими. Також завдання посібника — спонукати стейкхолдерів до дій за межами України, окреслюючи, як згадані межі можна адаптувати до інших країн, які перебувають в умовах кризи, надзвичайної ситуації чи збройного конфлікту.

Цей посібник є живим документом, який може змінюватися відповідно до контексту подій в Україні та за її межами.

Вступ

Коли в умовах кризи, надзвичайної ситуації та збройного конфлікту ми опиняємося на межі життя та смерті, безпечний доступ до перевіреної інформації може відігравати вирішальну роль. Як кризові протоколи для фізичного світу, протоколи для цифрової сфери здатні рятувати людські життя.

Платформи соціальних медіа стали критично важливими цифровими інфраструктурами для доступу до інформації та її поширення. Вони відіграють ключову роль у просуванні перевіреної інформації та протидії дезінформації й мові ворожнечі у конфліктах на кшталт повномасштабного вторгнення в Україну.

Для України платформи соціальних мереж є вкрай важливими під час повномасштабного вторгнення. 84 % українців використовують соціальні мережі як основне джерело новин, а для 42 % вони — єдиний канал отримання інформації[3].

З 2014-го, а особливо з 2021 року, Російська Федерація веде в Україні гібридну війну безпрецедентних масштабів, і тактика інформаційної війни в ній переплітається з реальними бойовими діями. Проте в контексті гібридної війни випадок з Україною — не єдиний. Навпаки, тактика інформаційної війни й іноземне втручання стають глобальною проблемою.

Повномасштабна війна серйозно посилила таргетованість і вразливість жінок, меншин і маргіналізованих груп, а також призвела до появи нових вразливих категорій. Зокрема, жінки, меншини й маргіналізовані групи стикаються з підвищеним ризиком насильства, переміщення й економічних труднощів.

Надзвичайні ситуації, кризи та збройні конфлікти, а також дезінформація та мова ворожнечі непропорційно впливають на жінок, меншини та маргіналізовані групи. Особи, що перебувають у центрі уваги громадськості, як-от політики, журналісти, жінки-військовослужбовці та активісти, мають

[3] Schafer, Bret et al., 2022.

високий ризик стати об'єктами онлайн-атак. Щоб не лише виявляти ризики, а й розробляти ефективні заходи щодо пом'якшення можливих наслідків, сприяючи захисту прав людини та демократичних цінностей, важливо узгоджувати зусилля та ініціативи з потребами й правами жінок, меншин і маргіналізованих груп.

Посібник з управління ризиками в контексті надзвичайних ситуацій, збройних конфліктів та криз є результатом співпраці багатьох стейкхолдерів: міжсекторної національної експертної робочої групи, до складу якої увійшли представники організацій громадянського суспільства, державних та регуляторних органів, наукових кіл та медіа; окремих місцевих, регіональних і міжнародних експертів, а також компаній.

Згадані в посібнику ризики визначені на основі досвіду України з лютого 2022-го до лютого 2025 року, водночас особливу увагу приділили періоду з лютого по грудень 2022 року.

Хід війни — мінливий, іноді щось змінюється щодня чи навіть щогодини. Тож заходи щодо зменшення ризиків мають бути гнучкими й такими, які можна легко адаптувати залежно від оцінювання ситуації. Управління ризиками та рекомендації щодо пом'якшення наслідків можуть функціонувати як життєво важливі інструменти, що дають змогу стейкхолдерами діяти швидко, особливо під час ескалації та загострення збройних конфліктів. Цей посібник орієнтується на принципи, викладені в «Рекомендаціях щодо управління цифровими платформами»[4] ЮНЕСКО та «Декларації принципів управління контентом і платформами під час кризи»[5] від Access Now, та відповідає їм, а також наголошує на важливості гнучкості, пропорційності та врахування місцевого контексту під час модерації контенту й поширення інформації.

З огляду на гібридний характер конфлікту, важливо враховувати високий ризик зовнішнього інформаційного впливу в регіоні та за його межами, який може посилюватися в періоди припинення вогню та після закінчення конфлікту. Тому індивідуальний підхід потрібен не лише для країн, безпосередньо залучених у конфлікт, але й для країн регіону та світу, яким він потенційно загрожує.

У контексті кризи, надзвичайної ситуації чи збройного конфлікту скорочення або відмова від фактчекінгу і співпраці з незалежними фактчекінговими організаціями може значно збільшити ризик дезінформації та мови ворожнечі, особливо в кризових регіонах. Досвід України засвідчив важливість систематичної співпраці між компаніями та надійними партнерами для захисту населення України від дезінформації та мови ворожнечі, а також для підтримки доступу до інформації. У нестабільному контексті будь-яке припинення або послаблення таких зусиль мало б негативні наслідки.

[4] UNESCO, 2023.
[5] Access Now, 2022.

Матриця ризиків на основі досвіду проживання збройного конфлікту в Україні

Матриця ризиків сформована крізь призму повномасштабної війни проти України, що визначає специфіку оцінювання ризиків. Вона ґрунтується на задокументованих випадках інформаційних атак, алгоритмічних помилок, маніпулювання контентом та використання соціальних платформ як інструментів гібридної війни.

Зазначені в матриці ризики виокремлені на основі досвіду України з лютого 2022-го до січня 2025 року, але особлива увага приділена періоду лютий — грудень 2022 року.

Місцева експертна робоча група визначила десять ключових ризиків (див. Табл) і класифікувала кожен як середній або високий. Ризики низького рівня не увійшли до таблиці. Під час різних фаз кризи можуть застосовувати різні підходи до пріоритезації ризиків.

Таблиця

Ключові ризики на соціальних платформах під час збройного конфлікту в Україні	**Вплив**	**Рівень пріоритетності**
1. Блокування та/або зменшення охоплення пов'язаного з війною контенту, який не порушує стандартів спільноти	Високий	Високий
2. Видалення контенту, який документує воєнні злочини	Високий	Середній
3. Використання ботів і фейкових акаунтів для поширення дезінформації та мови ворожнечі	Високий	Високий
4. Наявність неправдивого, оманливого та шкідливого контенту (зокрема згенерованого ШІ)	Високий	Високий
5. Неефективні інструменти для пошуку на платформах перевіреної інформації, пов'язаної з війною	Високий	Високий
6. Політики та практики модерації, що не враховують розуміння лінгвістичного, соціального, політичного, історичного та культурного контексту, зокрема щодо меншин, гендерних та маргіналізованих груп	Середній	Високий
7. Рекомендації користувачам шкідливого контенту, включно з дезінформацією та мовою ворожнечі	Високий	Високий
8. Брак можливості охоплювати користувачів на тимчасово окупованих територіях України	Високий	Високий
9. Зловживання комерційними інструментами в політичних та воєнних цілях, що порушують стандарти спільноти	Високий	Середній
10. Неадаптованість політик, стандартів та практик компаній до кризового контексту	Середній	Середній

Десять ключових рекомендацій на основі проживання досвіду збройного конфлікту в Україні

1. **Підтримувати обмін пов'язаним з війною контентом у суспільних інтересах відповідно до стандартів спільноти**
   Посилити сфокусованість на контенті, щоб гарантувати, що пов'язаний з війною контент, який відповідає стандартам спільноти, не буде видалений чи обмежене його охоплення.

2. **Зберігати контент, що документує воєнні злочини**
   Активізувати й ширше комунікувати зусилля, спрямовані на захист та збереження контенту, який документує воєнні злочини, щоб підтримати їхню реєстрацію.

3. **Боротися з дезінформацією від ботів та фейкових акаунтів**
   Докладати більше зусиль, щоб не допустити поширення дезінформації та мови ворожнечі за допомогою ботів і фейкових акаунтів.

4. **Видаляти неправдивий або шкідливий контент**
   Зосередитись на видаленні оманливого або шкідливого контенту, зокрема згенерованого ШІ, та контенту, націленого на вразливі групи — жінок, меншини й маргіналізовані спільноти.

5. **Забезпечувати доступ до перевіреної публічної інформації під час конфлікту**
   Гарантувати користувачам простий доступ до життєво важливої публічної інформації під час конфлікту.

6. **Ураховувати місцевий контекст під час рішень щодо модерації**
   Ретельніше враховувати в політиках модерації контекстуальні лінгвістичні, соціальні, політичні, історичні та культурні особливості, зокрема щодо меншин, гендерних і маргіналізованих груп.

7. **Зменшувати поширення шкідливого контенту через рекомендації**
   Зміцнити механізми, які запобігали б рекомендуванню користувачам шкідливого контенту, зокрема дезінформації та мови ворожнечі.

8. **Посилювати співпрацю з місцевими представниками**
   Сприяти співпраці з місцевими представниками, щоб забезпечити доступ до перевіреної інформації на тимчасово окупованих територіях.

9. **Запобігати зловживанням комерційними інструментами для шкоди**
   Докласти більше зусиль, аби гарантувати, що комерційні інструменти не використовуватимуться

в політичних та воєнних цілях, які порушують стандарти спільноти, вводять в оману або завдають шкоди іншими способами.

10. **Адаптувати політики компаній до кризових ситуацій**
    Додатково скоригувати політики та практики, щоб ефективно реагувати під час кризи, зокрема врахувати спричинені ними унікальні виклики.

Ці рекомендації визначили місцеві експерти, що перебували й працювали в українському контексті до та під час збройного конфлікту.

Кризовий протокол

Цей кризовий протокол містить рекомендації щодо того, як компанії можуть пом'якшити вплив дезінформації та мови ворожнечі з огляду на досвід проживання збройного конфлікту в Україні. Важливо регулярно оцінювати й переглядати ефективність і вплив згаданих у протоколі заходів, оскільки потреби та терміновість часто відрізняються під час різних фаз конфлікту. Наприклад, може варіюватися чисельність команд модераторів або вразливість певних груп.

Важливо: кроки, рекомендовані для початкових фаз, слід постійно переглядати й адаптувати в міру розвитку кризи.

До настання надзвичайної ситуації, кризи та збройного конфлікту: підготовка й запобігання

1. **Розробляти й регулярно переглядати кризові протоколи**
   - Розробити контекстно залежні протоколи, які можна активувати у випадку кризи. Регулярно оцінювати їх разом з місцевими експертами, щоб протоколи відповідали новим загрозам і були адаптовані до місцевого контексту, зокрема чутливі до ризиків і шкоди, пов'язаних з меншинами, гендерними й маргіналізованими групами. Протоколи можуть містити матрицю ризиків, зразок якої наведений вище.
   - Забезпечити виділення ресурсів для того, щоб персонал розумів контекст і міг запобігати ризикам і шкоді від дезінформації та мови ворожнечі, пов'язаних із кризою. Мова про всіх залучених осіб — від модераторів до тих, хто приймає рішення на рівні політики.

2. **Створити команду з управління кризою**
   - Зібрати спеціальну команду з управління кризою, до складу якої мають увійти внутрішні співробітники та місцеві стейкхолдери, які мають досвід реагування на кризи й глибоко обізнані з місцевим контекстом, зокрема лінгвістичними, соціальними, політичними, історичними, правовими та культурними аспектами регіону.
   - Забезпечити наявність підрозділів та механізмів, які спеціально відстежуватимуть, аналізуватимуть і впроваджуватимуть ініціативи з протидії гендерній дезінформації та мові ворожнечі;

захищатимуть жінок, меншини, маргіналізовані та вразливі групи, зокрема осіб, які перебувають у центрі уваги громадськості.

3. **Постійно переглядати політики, алгоритми та процеси модерації**
   - Залучати відповідних експертів, місцевих стейкхолдерів та представників постраждалих спільнот і вразливих груп, щоб регулярно переглядати реагування на надзвичайну ситуацію, кризу та збройний конфлікт, характер яких постійно змінюється, зокрема розвиток, тривалість та наслідки цих ситуацій.

4. **Співпрацювати з місцевими експертами та стейкхолдерами**
   - Ініціювати та зміцнювати співпрацю з місцевими організаціями, зокрема з надійними партнерами, фактчекерами, місцевими незалежними медіаорганізаціями, групами громадянського суспільства та організаціями, що спеціалізуються на надзвичайних ситуаціях, кризах і збройних конфліктах, щоб об'єднати зусилля з протидії дезінформації та мові ворожнечі відповідно до контексту. Наприклад, інформувати про політики модерації, переглядати кризові протоколи, заходи з моніторингу та оцінювання, а також сприяти розробленню ефективного реагування на дезінформацію та мову ворожнечі.

5. **Готувати верифіковані списки користувачів**
   - Разом з місцевими стейкхолдерами скласти списки верифікованих локальних користувачів, включно з фактчекерами, незалежними медіа та іншими стейкхолдерами, які мають досвід у боротьбі з дезінформацією та мовою ворожнечі та надають перевірену інформацію. Вони повинні забезпечити швидке оскарження для кола залучених учасників.

6. **Співпрацювати з фактчекерами та дослідниками**
   - Ставати партнерами та/або зміцнювати партнерство з місцевими, регіональними та глобальними фактчекерами й дослідниками, щоб створювати ефективні системи й канали обміну інформацією, знаннями та даними про події, пов'язані з дезінформацією та мовою ворожнечі.

7. **Підвищувати зусилля, спрямовані на забезпечення прозорості та підзвітності**
   - Забезпечити прозорість щодо того, як і чому певний контент видаляють чи позначають. Надавати регулярні звіти із зазначенням кількості видалених публікацій та обґрунтуванням цього.
   - Підвищити прозорість щодо рекламних інструментів та фінансування реклами, щоб гарантувати, що комерційні інструменти не стануть лазівкою для дезінформації та мови ворожнечі.
   - Розкрити, як компанії використовують інструменти ШІ для моніторингу контенту та які є щодо них обмеження, особливо в складних емоційно забарвлених ситуаціях на кшталт збройного конфлікту.
   - Інформувати користувачів про розвиток стратегій модерації контенту й оновлення, адаптовані до кризових ситуацій.

8. **Ініціювати співпрацю між платформами**
   - Разом з місцевими стейкхолдерами та незалежними дослідниками розробити ефективні шляхи комунікацій з іншими компаніями, щоб запобігати поширенню дезінформації та мови ворожнечі з однієї платформи соціальних медіа на іншу.

9. **Ініціювати зусилля зі збереження контенту, що документує воєнні злочини**
   - Визначити та розробити кризові протоколи з відповідними архівними стейкхолдерами, щоб підготувати й ініціювати заходи зі збереження контенту, що документує воєнні злочини.

10. **Адаптувати приватні політики та практики оброблення даних**
    - Усунути вразливості, які можуть наражати користувачів на загрозу стеження, таргетування або маніпуляцій, надаючи їм інструменти для захисту своїх даних та конфіденційності. Особливу увагу приділяти військовополоненим та їхнім мережам.

Під час надзвичайної ситуації, кризи та збройного конфлікту: реагування та захист у режимі реального часу

1. **Моніторити та виявляти загрози в реальному часі**
   - У режимі реального часу моніторити на своїх платформах події та тенденції, що стосуються пов'язаної з конфліктом дезінформації та мови ворожнечі, включно з гендерною дезінформацією та проявами штучної діяльності. Залучити до цього і автоматизовані системи, і людей-модераторів, а відповідну інформацію отримувати від фактчекерів, дослідників та місцевих стейкхолдерів.

2. **Співпрацювати з місцевими експертами для оновлень у реальному часі**
   - Залучати місцевих експертів, медіа, групи громадянського суспільства та стейкхолдерів, зокрема в зонах активного конфлікту та на тимчасово окупованих територіях, щоб гарантувати: рішення щодо модерації контенту є безпечними, відповідають контексту та захищають групи, вразливі до дезінформації та ворожнечі, й/або ті, що є їхніми ймовірними об'єктами.

3. **Взаємодіяти з надійними партнерами для швидкого фактчекінгу**
   - Зміцнювати співпрацю з надійними фактчекінговими організаціями, місцевими медіа та іншими сторонами, зацікавленими в перевіреній інформації, щоб забезпечити точну та швидку перевірку пов'язаних із кризою даних.

4. **Зміцнювати прозорість та підзвітність**
   - Публічно звітувати про свої дії, зокрема про рішення щодо модерації та зміни політики, перед місцевими стейкхолдерами та користувачами платформ, щоб завоювати довіру та поліпшити їхні знання щодо використання платформ.

5. **Надавати точну та безпечну інформацію, на місцевому рівні пов'язану з кризою, надзвичайною ситуацією чи збройним конфліктом**
    - Сприяти гуманітарній та безпековій інформації, що важлива для захищеності та добробуту місцевого населення.

6. **Підтримувати незалежних фактчекерів**
    - Продовжувати надавати технічну та фінансову підтримку незалежним фактчекерам, щоб забезпечити їхню неупередженість та об'єктивність під час конфлікту. Наголошувати на використанні відкритих методологій і прозорій звітності, щоб зміцнити довіру користувачів.

Після надзвичайної ситуації, кризи та збройного конфлікту: відновлення, оцінювання й довгостроковий моніторинг

1. **Оцінити й проаналізувати реагування на кризу**
    - Провести посткризове оцінювання, щоб визначити ефективність реагування на кризу, зокрема те, наскільки добре вдалося впоратися з дезінформацією та мовою ворожнечі. До проведення цього аналізу варто залучити місцевих, регіональних та глобальних експертів, медіа, фактчекерів та інших відповідних стейкхолдерів, щоб зібрати інформацію для реагування на майбутні кризи.

2. **Зміцнювати співпрацю для довгострокової стабільності**
    - Продовжити співпрацю з місцевими стейкхолдерами, групами громадянського суспільства, фактчекерами та незалежними дослідниками, щоб моніторити процес відновлення й запобігти повторному поширенню дезінформації та мови ворожнечі.

3. **Моніторити постконфліктну дезінформацію та мову ворожнечі**
    - Продовжити моніторинг, реагувати на поширення дезінформації та мови ворожнечі впродовж щонайменше двох років після закінчення безпосередньої кризи, щоб сприяти стабільності й підтримувати мирні процеси.

4. **Забезпечити прозорість і підзвітність**
    - Публічно звітувати про свої дії, включно з рішеннями щодо модерації контенту, видалення дезінформації, а також про те, як вплинула на ті чи інші рішення інформація від місцевих експертів. Така прозорість допомагає відновити довіру й демонструє відповідальність за роль платформи під час кризи.

5. **Постійно вдосконалювати кризові протоколи**
    - Удосконалити свої протоколи керування кризою на основі оцінювання конфлікту та його наслідків. Регулярно залучати місцевих експертів та стейкхолдерів до оновлення політик і практик для майбутніх кризових сценаріїв.

Рекомендовані українські організації для співпраці

Цей список невеликий за обсягом, проте він може стати стартовою точкою для подальшої співпраці та налагодження зв'язків із відповідними місцевими стейкхолдерами.

ЦЕДЕМ (Центр демократії та верховенства права)
Напрями: аналітика та експертиза з питань свободи слова, реформування медіазаконодавства, адвокаційні кампанії щодо прозорого регулювання онлайн-простору.

Інститут масової інформації (IMI)
Напрями: моніторинг та аналіз порушень прав журналістів, фактчекінг щодо дезінформації, навчання для представників медіа з питань безпеки та професійних стандартів.

Інтерньюз-Україна
Напрями: підготовка фахівців у галузі медіаграмотності та цифрової безпеки, проведення досліджень інформаційного простору, просування стандартів якісної журналістики та боротьба з дезінформацією.

Лабораторія цифрової безпеки
Напрями: консультації щодо захисту даних і конфіденційності, реагування на кіберзагрози та координація зусиль у кризових ситуаціях.

Міністерство цифрової трансформації України
Напрями: ініціювання нормативних та законодавчих змін у галузі цифрових технологій, надання консультацій щодо державних пріоритетів у галузі інформаційної безпеки, співпраця з великими технологічними компаніями задля формування стратегічних підходів до захисту даних і протидії дезінформації.

Ukrainian Archive
Напрями: фахове збирання, архівування та каталогізація даних, пов'язаних із воєнними злочинами та порушеннями прав людини; консультації щодо довготривалого зберігання, обробки та перевірки автентичності цифрових матеріалів; підтримка міжнародних розслідувань шляхом забезпечення доступу до оцифрованих доказів.

StopFake
Напрями: перевірка та спростування неправдивої інформації про події в Україні, аналіз кремлівської пропаганди, просвітницька діяльність щодо підвищення рівня медіаграмотності та критичного мислення в суспільстві.

VoxUkraine

Напрями: фактчекінг заяв політиків, бізнесменів, блогерів та інших публічних осіб; аналіз та спростування фейків у публічному дискурсі; надання аналітичних матеріалів щодо стратегій боротьби з дезінформацією.

Жінки в медіа

Напрями: аналіз гендерного балансу у сфері медіа, співпраця з жінками в журналістиці та медіаменеджменті, експертиза гендерно чутливої політики в медіаіндустрії, дослідження гендерних стереотипів у висвітленні новин.

Список літератури

Access Now. (2022, November 29). *Declaration of principles for content and platform governance in times of crisis.* https://www.accessnow.org/wp-content/uploads/2022/11/Declaration-of-principles-for-content-and-platform-governance-in-times-of-crisis.pdf

Alaphilippe, A., Machado, G., Miguel, R., & Poldi, F. (2022, September 27). *Doppelganger: Media clones serving Russian propaganda.* EU Disinfo Lab. https://www.disinfo.eu/wp-content/uploads/2022/09/Doppelganger-1.pdf

Aleksejeva N., Osadchuk, R., Gelava, S., Le Roux, J., Caniglia, M., Suárez Pérez, D., & Kann, A. (2022, September 27). *Russia-based Facebook operation targeted Europe with anti-Ukraine messaging.* Digital Forensic Research Lab (DFR Lab). https://medium.com/dfrlab/russia-based-facebook-operation-targeted-europe-with-anti-ukraine-messaging-389e32324d4b

Centre for Democracy and Rule of Law (CEDEM). (2024, January 25) *Recommendations for improving the moderation of Ukrainian content about Russia's armed aggression on Meta platforms.* https://cedem.org.ua/library/rekomendatsiyi-sotsmerezhi/

Centre for Strategic Communications and Information Security, & Centre for Democracy and Rule of Law (CEDEM). (2024, April 24). *Informational attacks in social networks: Research on Russian disinformation influence through advertising on Facebook.* https://spravdi.gov.ua/en/yak-rosiya-atakuye-ukrayinu-dezinformacziyeyu-cherez-reklamu-v-facebook-doslidzhennya-czentru-strategichnyh-komunikaczij/

Counter Disinformation Network. (2024, September 3). *Fool Me Once: Russian influence Operation Doppelganger continues on X and Facebook.* https://alliance4europe.eu/wp-content/uploads/2024/09/CDN-Report-%E2%80%93-Fool-Me-Once_-Russian-Influence-Operation-Doppelganger-Continues-on-X-and-Facebook-%E2%80%93-September-2024.pdf

CyberPeace Institute. (2024, October). *Report of Second Expert Meeting on Harms Methodology.* https://cyberpeaceinstitute.org/wp-content/uploads/2024/10/Second-Expert-Meeting-Harms-Methodology-2024.docx.pdf

CyberPeace Institute. (2023). *Cyber dimensions of the armed conflict in Ukraine - Quarterly analysis report Q3 July to September 2023.* https://cyberpeaceinstitute.org/wp-content/uploads/2023/12/Cyber-Dimensions_Ukraine-Q3-2023.pdf

European Commission. (2018). *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation.* https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation

EEAS. (2024, January). *2nd EEAS Report on Foreign Information Manipulation and Interference Threats: A Framework for Networked defence.* https://www.eeas.europa.eu/eeas/2nd-eeas-report-foreign-information-manipulation-and-interference-threats_en

EEAS. (2023, February). *1st EEAS Report on Foreign Information Manipulation and Interference Threats: Towards a Framework of Networked defence.* https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en

Global Network Initiative. (2017, May). *GNI Principles on Freedom of Expression and Privacy.* https://globalnetworkinitiative.org/wp-content/uploads/2018/04/GNI-Principles-on-Freedom-of-Expression-and-Privacy.pdf

Grippo, V. (2024, December 4). *Regulating content moderation on social media to safeguard freedom of expression.* Committee on Culture, Science, Education and Media, Council of Europe. https://rm.coe.int/as-cult-regulating-content-moderation-on-social-media-to-safeguard-fre/1680b2b162

Hearing Before the United States House of Representatives Committee on Energy and Commerce Subcommittees on Consumer Protection & Commerce and Communications & Technology (2021, March 21). *Testimony of Mark Zuckerberg, Facebook, Inc.* https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-U1.pdf

Kyrychenko, Y., Brik, T., van der Linden, S., & Roozenbeek, J. Social identity correlates of social media engagement before and after the 2022 Russian invasion of Ukraine. *Nature Communications 15, 8*127 (2024). https://doi.org/10.1038/s41467-024-52179-8

Mantas, H. (2021, May 13). *Sen. Mark Warner says he is embarrassed by congressional inaction on tech regulation.* Poynter.org. https://www.poynter.org/fact-checking/2021/sen-mark-warner-embarrassed-by-congressional-inaction-on-tech-regulation/

Porter, E., & Wood, T. J. (2021, September 10). *The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom.* PNAS. https://doi.org/10.1073/pnas.2104235118

Schafer, B., Benzoni, P., Koronska, K., Rogers, R., & Reyes, K. (2024, May 30). *Russian propaganda as a nesting doll: How RT layered the digital information environment.* German Marshall Fund. https://www.gmfus.org/sites/default/files/2024-05/Laundromat%20Paper.pdf

Semenyuta, I. (2023, June 26). *Moderation during the war: why social networks delete posts of Ukrainians.* Detector Media. https://ms.detector.media/sotsmerezhi/post/32269/2023-06-26-moderatsiya-pid-chas-viyny-za-shcho-sotsmerezhi-vydalyayut-dopysy-ukraintsiv/

Snopok, O. (2022, November 23). *"Potentially unacceptable." How the Russian war in Ukraine affects content moderation on social networks.* Detector Media. https://ms.detector.media/it-kompanii/post/30718/2022-11-23-potentsiyno-nepryynyatnyy-yak-rosiyska-viyna-v-ukraini-vplyvaie-na-moderatsiyu-kontentu-v-sotsmerezhakh/

Ukrainian Media and Communication Institute. (2023). *Media literacy for senior people (60+).* https://www.jta.com.ua/wp-content/uploads/2023/11/UMCI_MediaLiteracy_60_UA.pdf

UNESCO. (2023). *Guidelines for the governance of digital platforms.* https://unesdoc.unesco.org/ark:/48223/pf0000387339

Van Erkel, P. F. A., van Aelst, P., de Vreese, C. H., Hopmann, D. N., Matthes, J., Stanyer, J., & Corbu, N. (2024). When are fact-checks effective? An experimental study on the inclusion of the misinformation source and the source of fact-checks in 16 European countries. *Mass Communication and Society, 27(5), 851–876.* https://doi.org/10.1080/15205436.2024.2321542

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication, 37*(3), 350–375. https://doi.org/10.1080/10584609.2019.1668894

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