

## Масштаб проблеми блокування соцмережами українських медіа, блогерів та користувачів (дані співпраці ЦЕДЕМ з Meta, 24 лютого 2022 – 6 листопада 2023)

Центр демократії та верховенства права, як довірений партнер Meta, має повноваження приймати скарги від користувачів Facebook та Instagram щодо безпідставного блокування дописів та видалення облікових записів і подавати апеляції щодо цих дій до Meta. Завдяки звітам ЦЕДЕМ, Meta може аналізувати, чи були правильними рішення алгоритмів, та вдосконалювати свою роботу, а також змінювати рішення алгоритмів. Така діяльність вже показала позитивні результати: у період з 24 лютого 2022 р. по 6 листопада 2023 р. 68% апеляцій експертів ЦЕДЕМ були задоволені, і пости або облікові записи користувачів Facebook були розблоковані або відновлені.

Meta цього аналізу – виявлення найпоширеніших випадків, коли алгоритми соціальних медіа реагують негативно, необхідність та ефективність таких реакцій, а також визначення ефективності діяльності ЦЕДЕМ щодо оскарження випадків блокування/видалення дописів/акаунтів.

### Категорії звернень до служби підтримки Meta

1. **Видалені публікації чи коментарі:** звернення, які стосуються видалення контенту модераторами через певні порушення Стандартів спільноти Facebook. У випадках, представлених у звіті, найчастішими причинами видалення стають порушення правил про мову ворожнечі або згадки «небезпечних осіб чи організацій». Менш поширеними випадками є хибна інтерпретація сатири модерацією, яка зчитує використання жартів як серйозний опис або схвалення певного проблематичного явища (наприклад, сатиричне обернення тези російської пропаганди про «українців-нацистів» може бути сприйняте як пропаганда нацизму).
2. **Заблокований профіль:** звернення, які стосуються блокування або будь-яких інших обмежень діяльності сторінки на Facebook.

Причини таких блокувань найчастіше є тими ж, що і у випадку з видаленням публікацій і коментарів.

3. **Дезінформація:** звернення, які стосуються поширення в українському інфопросторі проблематичних наративів, які грають на руку інформаційній війні РФ. Мета таких повідомлень: створення панічних настроїв за рахунок використання тез на кшталт «військове керівництво України є не компетентним», «бойовий дух українських військових низький», «контрнаступ не вдасться» тощо.
4. **Попередження:** звернення, які стосуються проміжної ланки до видалення публікацій чи блокування профілю. Причини використання попереджень з боку Facebook зазвичай такі самі, як і у згаданих двох категорій, але така санкція застосовується в більш «м'яких» випадках.
5. **Злам:** звернення, які стосуються спроб відновити доступ до особистої чи публічної сторінки після зламу.
6. **Тіньовий бан:** звернення, які стосуються сторінок, для яких Facebook штучно знижує охоплення й популярність загалом.
7. **Доступ:** звернення, які стосуються відновлення доступу до сторінки внаслідок різних обставин, які не є зломом. Наприклад, проблем з отриманням доступу до профілю з нового пристрою.
8. **Росіяни:** звернення, які охоплюють російських військових чи публічних осіб, які активно підтримують РФ у війні.
9. **Шахрайство:** підозрілі «одноденні» сторінки, які імітують державні інституції чи міжнародні організації нібито для «виплат» населенню України. Інші прояви шахрайства включають сторінки, які імітують службу підтримки Meta та звертаються до користувачів під різними приводами для заволодіння персональними даними.
10. **Інше:** одиничні звернення, теми яких не охоплюють згадані вище категорії.

### **Категоризація статусів вирішення звернень ЦЕДЕМ до служби підтримки Meta**

**Вирішено:** служба підтримки розв'язала питання на користь сторони звернення.

**Частково вирішено:** служба підтримки частково розв'язала питання (наприклад, відновила тільки частину публікацій, видалення яких було оскаржене).

**Не вирішено:** служба підтримки вирішила питання, але не на користь сторони звернення.

**Невідомо:** служба підтримки або не розв'язала проблему, або не оновлювала статус звернення (наприклад, нічого не писала після повідомлення, яке приходить автоматично).



## Статистика вирішення звернень до служби підтримки Meta

### Категорія 1. Видалені публікації чи коментарі

Ця група скарг становить **21%** від загальної кількості. Найпопулярніший сценарій видалень стосується Стандартів спільноти Facebook у частині про мову ненависті (вживання авторами публікацій небажаних слів на кшталт «русня», «свинособаки», «москалі» та будь-якої іншої лексики, яку модератори платформи зчитує як розпалювання ворожнечі на національному підґрунті) або небезпечні особи чи організації (такими можуть вважатися, наприклад, бригада Національної гвардії України

«Азов» або очільник ОУН(б) Степан Бандера). Основна проблема взаємодії модераторів контенту з такими публікаціями та коментарями – хибна інтерпретація політичного контексту. Подібна інтерпретація часто призводить до випадків, коли з платформи прибирають навіть сатиричний контент, який ШІ або модератори можуть сприймати за серйозний. Розв'язання питань у цій категорії – 9 днів. **Розподіл за статусом розв'язання скарг:** «Вирішено» – 43%; «Частково вирішено» – 8%; «Не вирішено» – 39%; «Невідомо» – 10%.

- **Успіх у розв'язанні скарг у категорії «Видалені публікації чи коментарі»**



**Середній час розв'язання скарги в категорії «Видалені публікації чи коментарі»: 9 днів.**

- **Приклади скарг у категорії**

### Кейс 1

**Коли:** серпень 2023 р.

**Сутність:** Видалення сатиричного посту, який містив скріншот інструкції для військовослужбових ЗС РФ зі здачі в полон. В інструкції вказано використати гранату для здійснення самогубства. Пост видалено як заклик до навмисного пошкодження свого тіла.

**Статус:** не вирішено. Служба підтримки Meta вказала, що публікація «дійсно порушує Стандарти спільноти» та не може бути відновленою на платформі.

**Час вирішення:** 0 днів.

## Кейс 2

**Коли:** вересень 2023 р.

**Сутність:** публікація регіонального медіа, яка містила символіку батальйону “Азов”. Публікацію видалено.

**Статус:** не вирішено. Служба підтримки Meta вказала, що публікація «дійсно порушує Стандарти спільноти» та не може бути відновленою на платформі.

**Час вирішення:** 0 днів.

## Кейс 3

**Коли:** серпень 2023 р.

**Сутність:** публікація зі згадкою 10 ОГШБр “Едельвейс” видалена. Імовірно, алгоритми соцмережі сприйняли цей пост як поширення нацизму, адже у складі Вермахту в часи Другої світової війни існував підрозділ з аналогічною назвою - 1 гірськопіхотна дивізія “Едельвейс”.

**Статус:** вирішено. Публікацію відновлено.

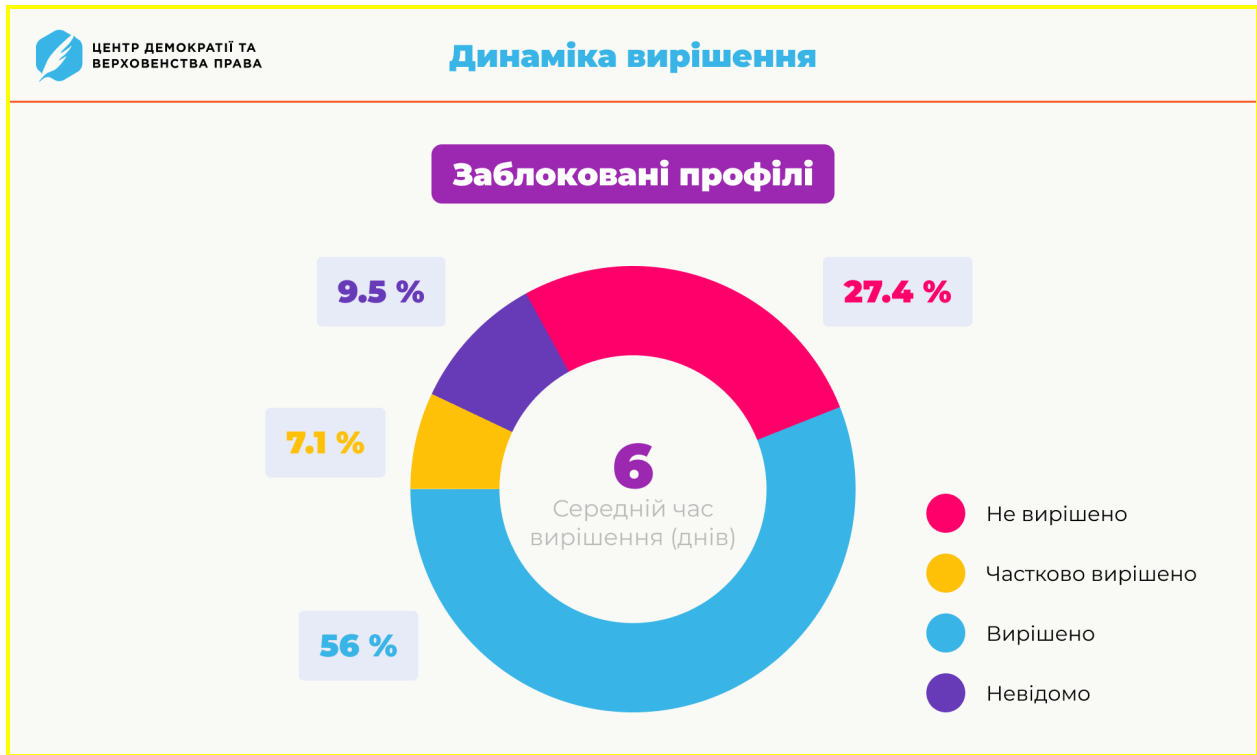
**Час вирішення:** 1 день.

## Категорія 2. Заблоковані профілі

Заблоковані профілі – категорія скарг у службу підтримки Meta, яка становить **14%** від загальної кількості скарг. Зазвичай профілі блокують з причин, подібних до видалення публікацій: переважно це інтерпретація

контенту як «мова ненависті» або згадка авторами «небезпечних» осіб чи організацій, але, на відміну від кейсів видалення публікацій, Facebook не вказує на пряму причину блокування сторінок. **Динаміка розв'язання скарг у цій категорії:** «Вирішено» – 56%; «Частково вирішено» – 7%; «Не вирішено» – 27%; «Невідомо» – 10%. Середній час реагування та вирішення звернень – 6 днів.

- **Успіх у розв'язанні скарг у категорії «Заблоковані профілі»**



**Середній час розв'язання скарги в категорії «Заблоковані профілі»: 6 днів.**

- **Приклади скарг у категорії «Заблоковані профілі»**

**Кейс 1**

**Коли:** серпень 2023 р.

**Сутність:** спільноту заблокували через публікації про нібито “нелегальний продаж” та “нанесення шкоди” тваринам, хоча оригінальні публікації були оголошеннями з пошуку зниклих тварин у Києві.

**Статус:** не вирішено. Служба підтримки вказала, що сторінка “порушувала Стандарти спільноти”.

**Час вирішення:** 1 день.

## Кейс 2

**Коли:** вересень 2023 р.

**Сутність:** без пояснення причини заблокували сторінку онлайн-магазину колекційних монет.

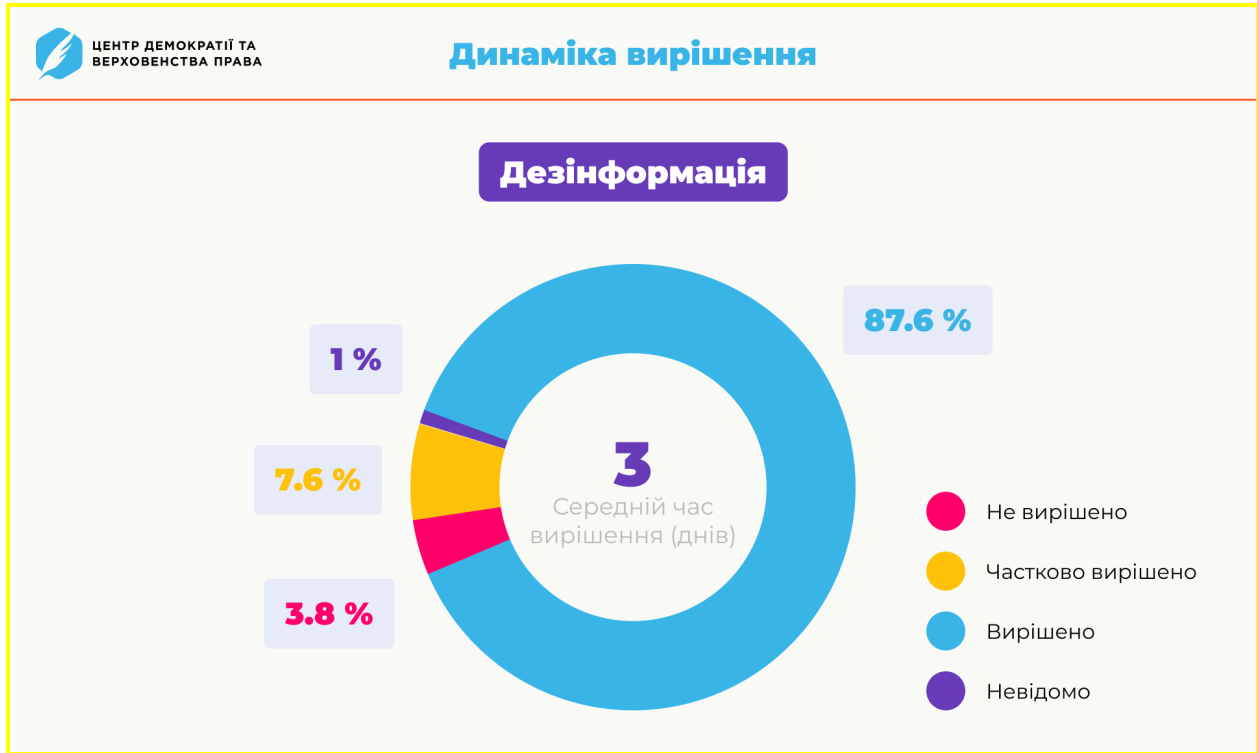
**Статус:** вирішено. Сторінку відновлено.

**Час вирішення:** 1 день.

## Категорія 3. Дезінформація

Поширення російської дезінформації на Facebook – виклик з перших днів повномасштабного вторгнення, але її суттєва активізація припала на весну 2023 року. Частка звернень стосовно дезінформації становить **17%**. Ця дезінформація поширюється через акаунти-«одноденки», які мають не пов’язані з війною назви (наприклад, *Colorful Nails SRD* або *Sensu*). Контент переважно артикулює такі теми: контрнаступ Збройних Сил України (акцент на проблемах із західною військовою технікою, некомпетентності військового керівництва тощо), мобілізація (а саме її опис як засобу підкорення та «знищення» населення), поширення неправдивих повідомлень для спричинення панічних настроїв (уявні хімічні атаки на далекі від фронту регіони України, відновлення режимів відключення світла) тощо. Успіх у протидії дезінформації має вдалий та системний характер: 88% скарг було задоволено, 8% – частково задоволено і лише 4% було відхилено. Час розв’язання таких скарг також суттєво перевищує середній показник – служба підтримки реагує на них в середньому за 3 дні (протягом літа й осені 2023 р. цей показник покращується: почастишали випадки, коли дезінформаційні кампанії видаляються в день звернення до служби підтримки Meta).

- **Успіх у розв'язанні скарг у категорії «Дезінформація»**



**Середній час розв'язання скарги в категорії: 3 дні.**

- **Приклади скарг у категорії «Дезінформація»**

**Кейс 1**

**Коли:** липень-листопад 2023

**Сутність:** майже всі звернення щодо дезінформації стосуються виявлення цілих мереж схожих сторінок, які поширюють однакову дезінформацію. Це можуть бути сотні сторінок в межах одного звернення.

**Статус:** майже 90% таких випадків видаляються.

**Час вирішення:** коливається від 1 до 3 днів.

**Кейс 2**

**Коли:** липень 2023.



**Сутність:** сторінка, яка систематично поширює російські наративи. Серед прикладів: “США здійснили удар по Кримському мосту”, використання емодзі свиней для позначення українців, “біолабораторії в Україні” тощо.

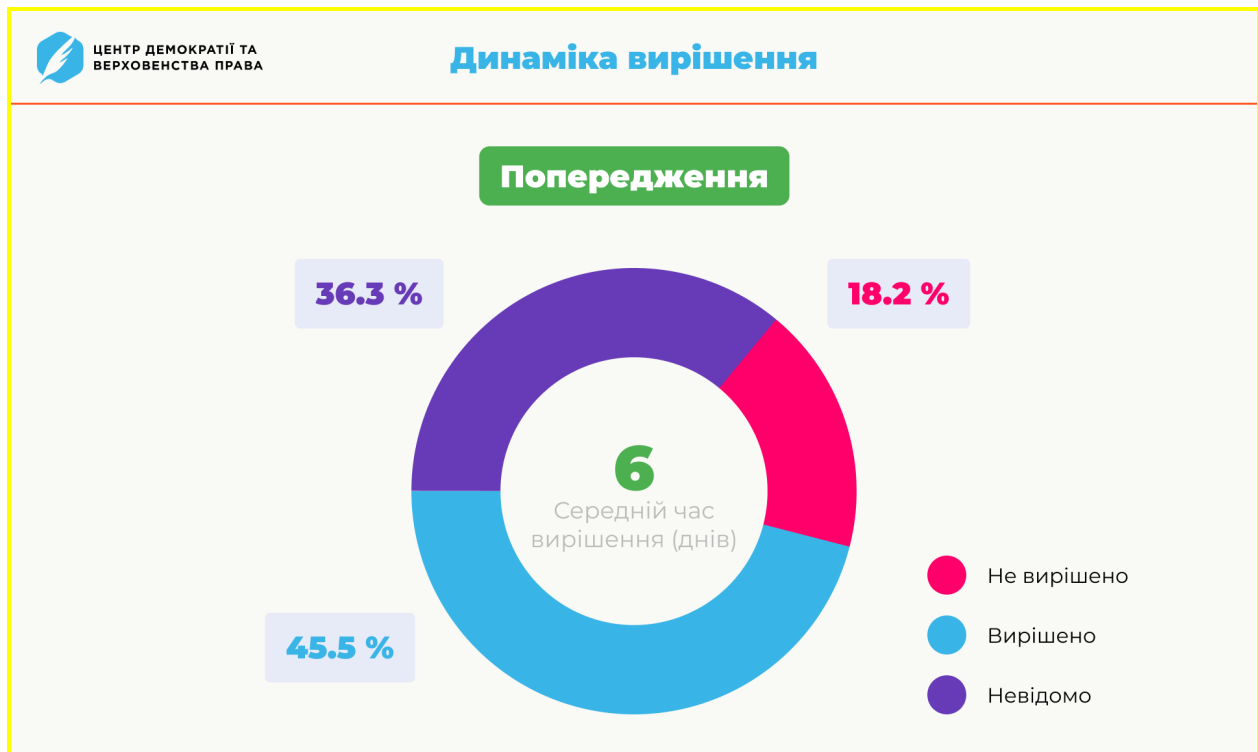
**Статус:** не вирішено. Служба підтримки не знайшла порушень Стандартів спільноти в такому контенті.

**Час вирішення:** 1 день.

#### Категорія 4. Попередження

У цю категорію звернень потрапляють скарги, які стосуються обмежень у кількості дописів, які може опублікувати сторінка; попереджень про потенційний блок сторінки у випадку повторного порушення Стандартів спільноти; штучного зниження охоплення сторінки тощо. Частка таких звернень становить **4%**. Служба підтримки часто не відповідає на звернення, які стосуються попереджень – таких випадків було 37%. У 45% ситуацію було розв’язано позитивно, а на 18% звернень служба підтримки відповіла негативно. Середній час вирішення проблем у цій категорії становить 6 днів.

- **Успіх у розв’язанні скарг у категорії «Попередження»**



**Середній час розв'язання скарги в категорії: 6 днів.**

- **Приклади скарг у категорії «Попередження»**

### **Кейс 1**

**Коли:** вересень 2023.

**Сутність:** журналіст мав санкції на акаунті, пов'язані з відображенням його публікацій, які пройшли ще в серпні. Тим не менше, у вересні він продовжував спостерігати схожі патерни, хоча формально на його сторінці вже не було порушень і Facebook не відображав санкцій на цьому акаунті.

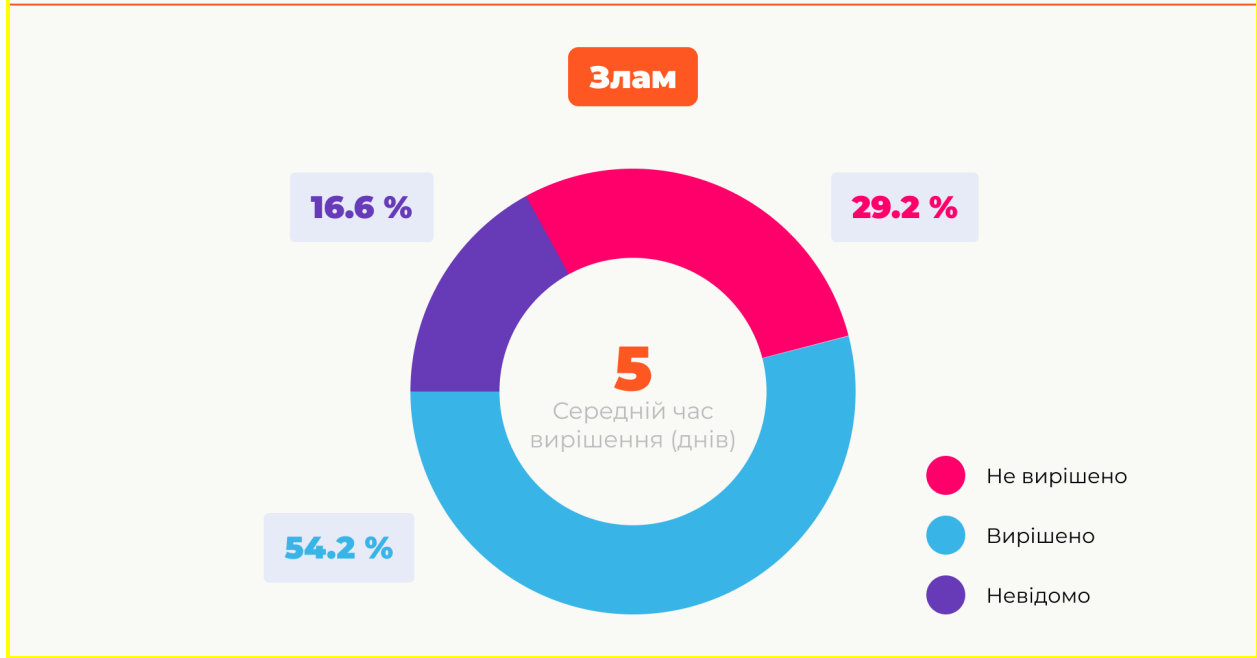
**Статус:** не вирішено. Служба підтримки переслала позивачів на іншу інстанцію - так звані [self reporting tools](#).

**Час вирішення:** 15 днів.

## **Категорія 5. Злам**

Ця категорія стосується звернень до служби підтримки Meta з проханням відновити доступ до профілів, які були зламані. Відсоток звернень становить всього **4%**. Частка вирішених проблем становить 54%, невирішених – 29%, а 17% звернень мають невідомий сценарій розвитку. Середній час реагування служби підтримки на такий тип звернень – 5 днів.

- **Успіх у розв'язанні скарг у категорії «Злам»**



**Середній час розв'язання скарги в категорії: 5 днів.**

- **Приклади скарг у категорії «Злам»**

### Кейс 1

**Коли:** липень 2023.

**Сутність:** сторінку арт-директора одного з українських фестивалів зламали та змінили на ній ім'я.

**Статус:** вирішено. Служба підтримки звернулася до власника через електронну пошту та сприяла відновленню доступу до сторінки.

**Час вирішення:** 2 дні.

### Кейс 2

**Коли:** серпень 2023.

**Сутність:** рекламний кабінет підприємця зламали та розповсюджували через нього реклами з інших сторінок, не пов'язаних з ним. Він заблокував ці сторінки, але як результат мав заборгованість за

поширення реклам, що спричинило блокування його рекламного кабінету.

**Статус:** не вирішено. Служба підтримки переслала позивачів на іншу інстанцію - так звані [self reporting tools](#).

**Час вирішення:** 26 днів.

### Категорія 6. Тіньовий бан

Ця категорія переважно містить звернення до служби підтримки, які стосуються зменшення охоплення публікацій через порушення Стандартів спільноти. Частка цієї категорії становить **1%**. 50% звернень були успішно розв'язані, 17% – не вирішені, а 33% залишилися без відповіді або мають невідомий сценарій розвитку. Середній час вирішення питання у цій категорії становить 12 днів.

- **Успіх у розв'язанні скарг у категорії «Тіньовий бан»**



**Середній час розв'язання скарги в категорії: 12 днів.**

- **Приклади скарг у категорії «Тіньовий бан»**

## Кейс 2

**Коли:** липень 2023.

**Сутність:** журналіст опублікував відео з двома жінками, які б'ються одна з одною в місті Ярославль у РФ. Відео він прокоментував у сатиричній манері реплікою “Курорт на росії. Так всьо уютна, па-дамашньому! Прієзжайте в Ярославль”. Публікація була переміщена нижче у стрічці, тобто у неї впали потенційні охоплення.

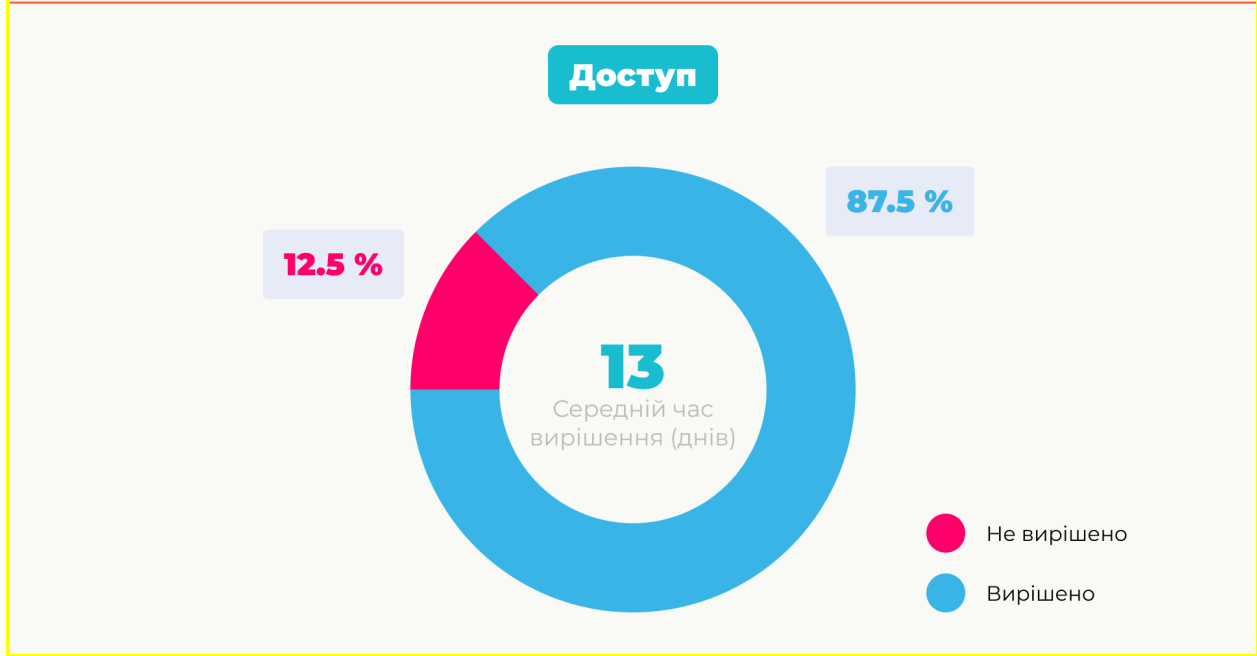
**Статус:** невідомо. Служба підтримки не відповідала після стандартного автоматичного повідомлення.

**Час вирішення:** відсутній.

### Категорія 7. Доступ

Ця категорія стосується втрати доступу до особистої чи публічної сторінки та спроб його відновити або підозрілої активності на сторінці з незвичної локації. Це одна з найменш чисельних категорій, вона становить **1%** від усіх звернень. 88% звернень були розв'язані, 12% - не вирішені. Середній час реагування становить 13 днів. Це одна з «найповільніших» категорій, але таке явище пов'язане з тим, що у подібних кейсах доволі багато часу йде на з'ясування справжніх власників сторінки.

- **Успіх у розв'язанні скарг у категорії «Доступ»**



**Середній час розв'язання скарги в категорії: 13 днів.**

- **Приклади скарг у категорії «Доступ»**

### Кейс 1

**Коли:** вересень 2023.

**Сутність:** користувач не зміг отримати доступ до профілю, адже мав двофакторну аутентифікацію та видалив сторонній додаток з нею з телефону. Через це він не міг зайти на сторінку навіть після первинного звернення до Служби підтримки.

**Статус:** вирішено. Служба підтримки звернулася до власника сторінки та сприяла відновленню доступу до акаунту.

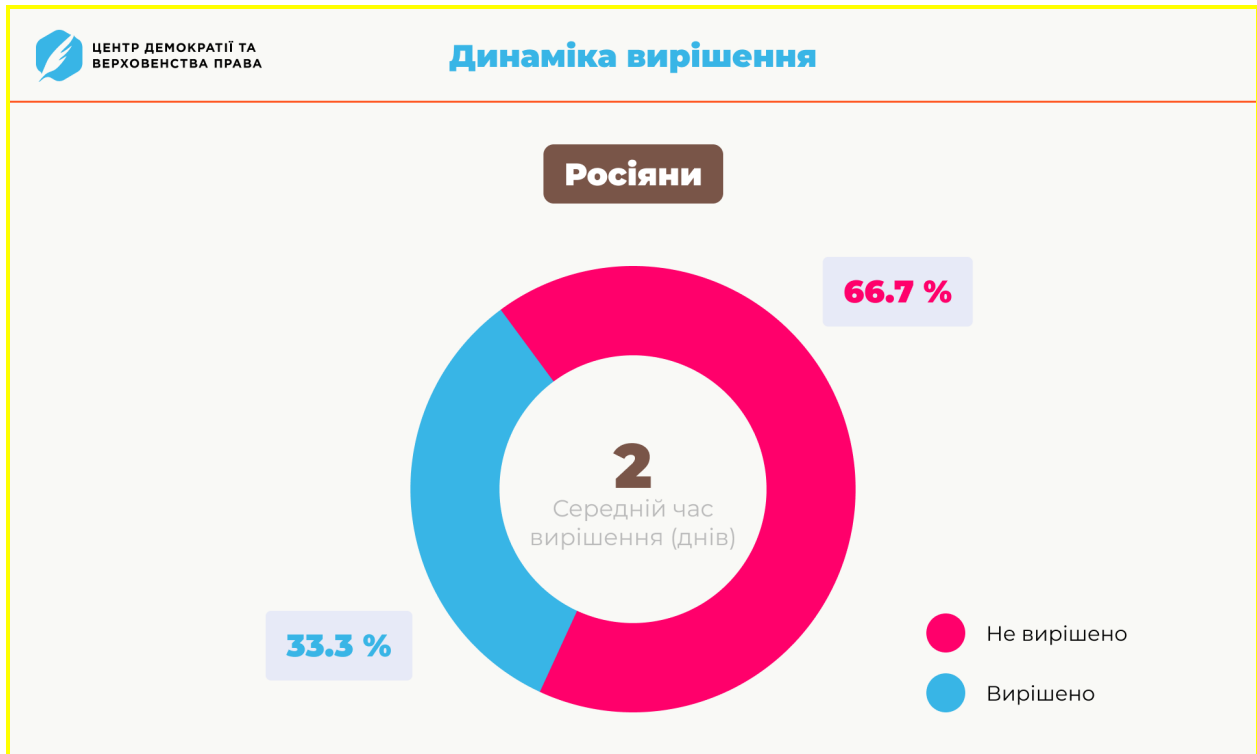
**Час вирішення:** 1 день.

### Категорія 8. Росіяни

Ця категорія звернень є найменш чисельною з усіх, вона становить **0.5%** звернень від загальної кількості. Тим не менше, вона варта виділення в

окремий розділ, адже дає змогу зрозуміти взаємодію Facebook з росіянами, які підтримують війну проти України. Звернення тут стосуються інтернет-сторінок російських військових і публічних осіб. Успіх у цій категорії: 33% вдалих звернень і 67% – невдалих. Середній час реагування – 2 дні.

- **Успіх у розв'язанні скарг у категорії «Росіяни»**



**Середній час розв'язання скарги в категорії: 2 дні.**

- **Приклади скарг у категорії «Росіяни»**

**Кейс 1**

**Коли:** вересень 2023.

**Сутність:** російська користувачка опублікувала допис, в якому відповідає на питання “Кто такой русский человек?” Допис містить елементи “м’якої” пропаганди, адже вказує на те, що “справжній росіянин” - це той, хто “пам’ятає історію та шанує пам’ять предків”; той, у кого “в крові допомагати та захищати” та інші подібні тези.

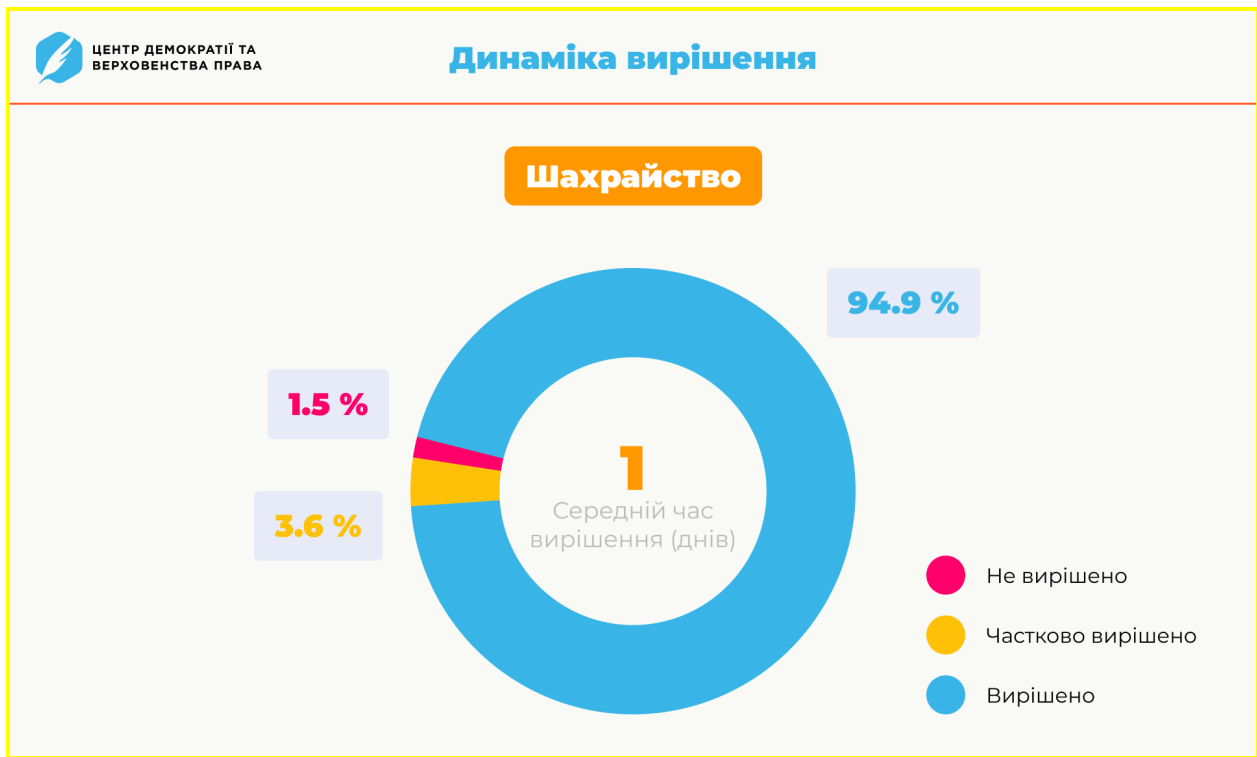
**Статус:** не вирішено. Служба підтримки не знайшла ознак порушення Стандартів спільноти.

**Час вирішення:** 0 днів.

### Категорія 9. Шахрайство

Це нова категорія відносно минулого звіту. Вона стосується активізації шахрайських сторінок протягом літа-осені 2023 року та за цей період стала найпоширенішою серед усіх категорій - **23%** від загальної кількості. Зазвичай вони мімікрують під офіційні сторінки державних інституцій або міжнародних організацій, які нібито здійснюють «виплати» для цивільного населення, або видають себе за службу підтримки Meta та пишуть користувачам з приводу вигаданих проблем, пов'язаних з акаунтами на Facebook (втрата доступу, видалення акаунту через певні «порушення» тощо). Розподіл успіху в категорії: 95% вдалих звернень, 4% - частково вдалі, 1% - невдалі. Середній час розв'язання скарги у категорії становить менше одного дня.

- **Успіх у розв'язанні скарг у категорії «Шахрайство»**





**Середній час розв'язання скарги в категорії: <1 день.**

- **Приклади скарг у категорії «Шахрайство»**

#### **Кейс 1**

**Коли:** липень-листопад 2023.

**Сутність:** у більшості випадків скарги на такі сторінки містять сотні прикладів шахрайських профілей з контентом, згаданим вище.

**Статус:** майже всі такі звернення вирішуються вдало.

**Час вирішення:** становить в середньому менше одного дня.

#### **Кейс 2**

**Коли:** липень-листопад 2023.

**Сутність:** деякі кейси містять змішану реакцію Служби підтримки. Сторінки можуть мати аналогічний контент, але платформа блокує вибірково та може залишити декілька відсотків сторінок з “пакету” звернення незаблокованими.

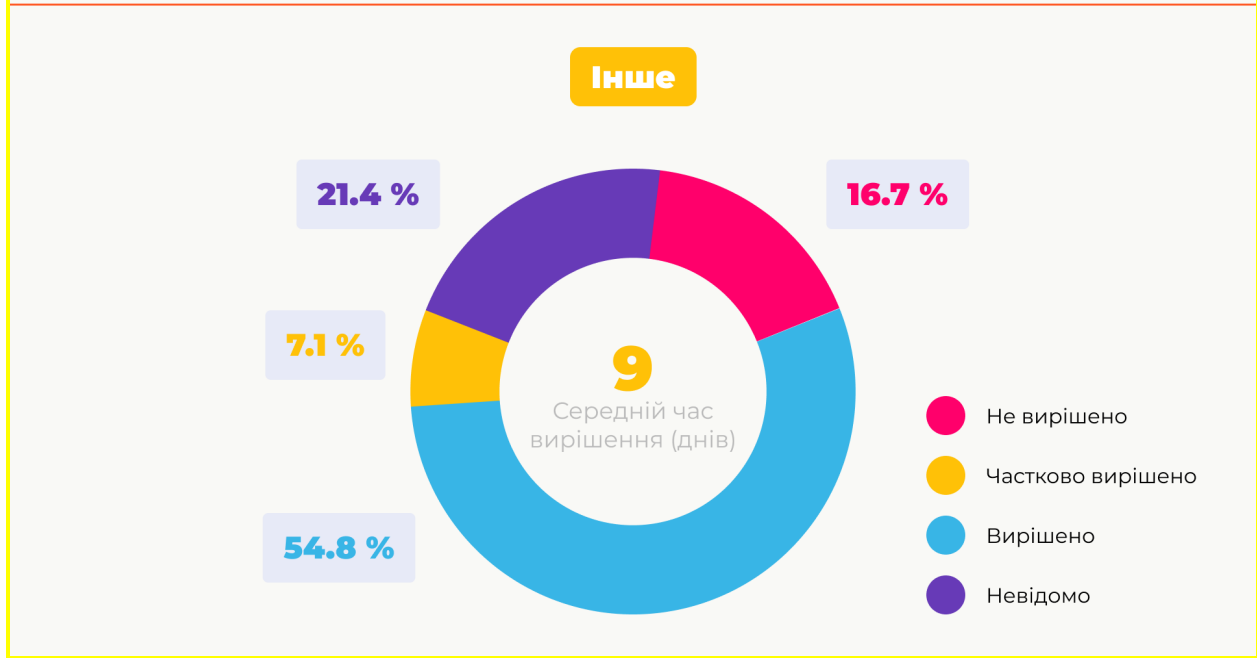
**Статус:** частково вирішено.

**Час вирішення:** 3 дні в середньому на підкатегорію “частково вирішених” звернень.

### **Категорія 10. Інші**

Категорія «Інші» містить одиничні звернення, які не вдалося віднести до жодної з категорій. Таких звернень всього **7%**. Вдале вирішення таких звернень має показник 55%, невдале – 17%, частково вдале – 7%, невідоме – 21%. Середній час вирішення скарг у цій категорії становить 10 днів.

- **Успіх у розв'язанні скарг у категорії «Інші»**



**Середній час розв'язання скарги в категорії: 10 днів.**

- **Приклади скарг у категорії «Інші»**

### Кейс 1

**Коли:** жовтень 2023.

**Сутність:** сторінка видавала себе за Валерія Залужного для реклами анонімного Telegram-каналу з новинами.

**Статус:** не вирішено. Служба підтримки не знайшла порушень Стандартів спільноти на сторінці. Зараз ця сторінка оперує під оновленою назвою.

**Час вирішення:** 0 днів.

### Висновки

**Загалом, з початку повномасштабного вторгнення по 6 листопада 2023 р. ЦЕДЕМ вдалося вирішити 68% звернень користувачів Facebook та Instagram.** Якщо говорити суто про відрізок від дати останнього спостереження (30 червня 2023 р.) до 6 листопада р., то

статистика ще краща – 78% вдало розв'язаних кейсів. **У категорії «Шахрайство» та «Дезінформація» ЦЕДЕМ та служба підтримки Meta співпрацюють найбільш вдало – 95% і 88% успіху відповідно.** Також, аналогічний рівень успіху (88%) і в категорії «Доступ» - служба підтримки активно сприяє відновленню контролю над втраченими сторінками. У випадку з шахрайськими сторінками та джерелами дезінформації помітна й швидка реакція – часто такі сторінки можуть видалятися модераторами платформи через декілька годин після подання скарги. Більше того, порівняно з 2022 р. суттєво пришвидшився процес вирішення скарг – 2 дні за період 30 червня – 6 листопада 2023 р. порівняно з 8.5 днів за 24 лютого 2022 р. – 30 червня 2023 р.. **За весь період співпраці з Meta середній час вирішення скарг становить трохи менше 6 днів.**

Як і за [минулий період](#) спостереження, у цей раз **найбільш «невдалою» категорією є «Видалені пости/коментарі» - 35% невдалих розв'язань кейсів** за період 30 червня – 6 листопада 2023 р. (39% за весь період). Причини лишаються такими ж: видалення сатиричних публікацій, хибна інтерпретація сталих виразів чи фразеологізмів та проблеми з розрізненням тональності публікацій, якщо йдеться про «небезпечних осіб чи організацій».