Дослідницький центр «Прізма»
Prism Research Center

Дослідницька записка

Вимірювання узгодженості щодо «придатності до дій»

Міжекспертна узгодженість у рецензуванні повідомлень

2025

Це видання: листопад 2026 р.

Дослідницька записка

Вимірювання узгодженості щодо «придатності до дій»

Міжекспертна узгодженість у рецензуванні повідомлень

Автор	Дослідницький центр «Прізма»
Опубліковано	2025
Це видання	Листопад 2026 р.
Період дослідження	Кінець 2024 – середина 2025 року
Ліцензія	Вільно читати, завантажувати й використовувати повторно. © Дослідницький центр «Прізма».
Контакти	research@prismresearch.org.ua · prismresearch.org.ua

Рекомендоване цитування

> Дослідницький центр «Прізма» (2025). *Вимірювання узгодженості щодо «придатності до дій»: міжекспертна узгодженість у рецензуванні повідомлень.* Дослідницька записка. Дослідницький центр «Прізма». Видання листопада 2026 р. https://prismresearch.org.ua/uk/publications/

Резюме

Будь-який метод тріажу публічних повідомлень навчається й оцінюється на основі людських позначок того, що є «придатним до дій». У цій записці ми з'ясовуємо, наскільки узгоджено двоє навчених аналітиків можуть ставити таку позначку і що з цього випливає. Ми відібрали 200 повідомлень із надходжень служби моніторингу, стратифікованих за шістьма категоріями повідомлень служби, і двоє аналітиків незалежно позначили кожне з них. Узгодженість становила κ Коена = 0,634 — суттєва узгодженість за загальноприйнятими мірками. Проста узгодженість (94%) була вищою за κ, оскільки кожен аналітик позначив як придатні до дій лише близько 9% повідомлень, тож значна частина збігів щодо «непридатних до дій» очікувана випадково. Більшість розбіжностей припала на граничні випадки: чутки, які згодом можуть підтвердитися, застарілі матеріали та повідомлення, що виходять за межі тематики, але дотичні до неї. Практичний наслідок — стеля: жодна модель тріажу не може бути надійнішою за позначки, на яких вона навчається.

κ = 0,634	200	9%
κ Коена: суттєва узгодженість за загальноприйнятими мірками	200 повідомлень, два аналітики: кожне позначено незалежно	Позначено як придатні до дій кожним аналітиком (6% після розв'язання розбіжностей)

Зміст

1 Питання
2 Метод
3 Вимірювання узгодженості
4 Результати
5 Узгодженість за категоріями повідомлень
6 У чому аналітики розійшлися
7 Наслідки для тріажу
8 Обмеження
9 Терміни

1 Питання

Наскільки узгоджено двоє навчених аналітиків можуть визначати, що робить публічне повідомлення придатним до дій, і що це означає для будь-якого методу тріажу, побудованого на таких позначках?

Це питання важливе, бо «придатність до дій» не є властивістю повідомлення, яку можна спостерігати безпосередньо. Це судження аналітика, і від нього залежить кожен наступний крок: позначки визначають, що саме має передбачати модель тріажу, і водночас є мірилом, за яким цю модель оцінюють. Якщо двоє уважних людей часто не погоджуються щодо позначки, частина будь-якої виміряної помилки моделі насправді є розбіжністю між людьми.

2 Метод

Вибірка	200 повідомлень із надходжень служби моніторингу, стратифікованих за шістьма стандартними категоріями повідомлень служби, з дотриманням регіонального балансу в межах кожної страти, наскільки дозволяли надходження. Повідомлень за категоріями: безпекові інциденти — 48; інфраструктура й логістика — 36; гуманітарні потреби — 40; офіційні заяви — 28; економічні умови — 24; неперевірені повідомлення — 24.
Експерти	Двоє навчених аналітиків.
Завдання	Кожен аналітик позначив кожне повідомлення як придатне або непридатне до дій за визначенням з інструкції: «Повідомлення є придатним до дій, якщо його огляд призводить щонайменше до одного з таких зафіксованих результатів: включення до щотижневого зведення про ситуацію; ескалації координатору служби для подальших дій; передачі партнерам спостереження щодо логістики чи доступу; або виправлення раніше поширеного твердження. Дублікати, неперевірені чутки, повідомлення поза тематикою та застарілі повідомлення не є придатними до дій». Експерти також отримали по одному розібраному прикладу для кожного типу результату з позначкою «Умовний приклад».
Незалежність	Кожен аналітик ставив позначки, не бачачи ні позначок іншого, ні жодних результатів моделі.
Розв’язання розбіжностей	Розбіжності розв’язував координатор служби, щоб сформувати робочий набір позначок.
Період дослідження	Кінець 2024 – середина 2025 року.

Стратифікація за категорією з регіональним балансом у межах кожної страти означає, що показник узгодженості описує весь набір матеріалів служби, а не лише найпоширеніші типи повідомлень. Оскільки аналітики не бачили позначок одне одного та результатів моделі, виміряна узгодженість — це узгодженість між незалежними судженнями, а не збіг, зумовлений орієнтацією на спільну підказку.

3 Вимірювання узгодженості

Найпростіший показник — проста відсоткова узгодженість: частка повідомлень, яким обидва аналітики поставили однакову позначку. Її легко зрозуміти, але вона може вводити в оману, коли одна позначка трапляється значно частіше за іншу. Якщо майже всі повідомлення непридатні до дій, двоє аналітиків, які позначають майже все як «непридатне до дій», здебільшого збігатимуться, навіть якщо рідко погоджуються щодо тих повідомлень, які справді важливі.

к Коена це виправляє. Він порівнює спостережену узгодженість з узгодженістю, очікуваною випадково з огляду на те, як часто кожен аналітик ставив кожну позначку:

$$\text{к} = (\text{спостережена узгодженість} - \text{випадкова узгодженість}) \div (1 - \text{випадкова узгодженість})$$

к, що дорівнює 1, означає повну узгодженість; 0 — узгодженість не більшу, ніж випадкова. За поширеною умовністю (Landis and Koch, 1977) значення від 0,41 до 0,60 вважають помірною, а від 0,61 до 0,80 — суттєвою узгодженістю. Ці межі є емпіричним правилом, а не стандартом, але дають спільну точку відліку.

4 Результати

Таблиця 1. Узгодженість двох аналітиків щодо «придатності до дій» (200 повідомлень, два аналітики)

Показник	Значення
Повідомлень, позначених обома аналітиками	200
Обидва аналітики: придатне до дій	12
Лише перший аналітик: придатне до дій	6
Лише другий аналітик: придатне до дій	6
Обидва аналітики: непридатне до дій	176
Проста відсоткова узгодженість	94% (188 з 200)
к Коена	0,634
Позначено як придатні до дій кожним аналітиком	9% (18 з 200)
Придатні до дій після розв’язання розбіжностей	6% (12 з 200)

Проста узгодженість була вищою за к. Такий розрив очікуваний, коли базовий рівень придатності до дій низький: випадкова узгодженість щодо «непридатних до дій» висока, і к її враховує.

На цій вибірці проста узгодженість становила 94%, а к Коена = 0,634 — суттєва узгодженість за загальноприйнятими мірками. 12 розбіжностей розподілилися порівну: шість повідомлень як придатні до дій позначив лише перший аналітик, і шість — лише другий. Навчені аналітики, працюючи незалежно, погоджувалися щодо того, які повідомлення придатні до дій, значно частіше, ніж випадково, але не завжди.

5 Узгодженість за категоріями повідомлень

Загальний показник к підсумовує всю вибірку. Щоб побачити, де саме аналітики погоджуються, а де розходяться, узгодженість можна розрахувати окремо для кожної категорії повідомлень.

Таблиця 2. Узгодженість за категоріями повідомлень

Категорія повідомлень	Повідомлень	Придатних до дій (після розв'язання розбіжностей)	Узгодженість
Безпекові інциденти	48	3	46/48 (95,8%)
Інфраструктура й логістика	36	2	34/36 (94,4%)
Гуманітарні потреби	40	2	38/40 (95,0%)
Офіційні заяви	28	2	26/28 (92,9%)
Економічні умови	24	1	23/24 (95,8%)
Неперевірені повідомлення	24	2	21/24 (87,5%)
Усього	**200**	**12**	**188/200 (94,0%)**

Шість стандартних категорій повідомлень служби. Узгодженість — це частка повідомлень, яким обидва експерти поставили однакову позначку. Відсотки округлено до одного знака після коми; кількості точні. к = 0,634 розраховано на об'єднаній вибірці (n = 200); узгодженість за категоріями наведено описово, оскільки групи замалі для стабільних оцінок к на рівні категорій.

Як читати таблицю. Проста узгодженість висока в кожній категорії, бо більшість повідомлень не є придатними до дій; практична складність полягає в характері розбіжностей, які зосереджені в неперевірених повідомленнях і граничних випадках. Кожна категорія містить лише частину з 200 повідомлень, тож відмінності між категоріями ґрунтуються на кількох розбіжностях і є орієнтовними.

Повідомлення з категорії «Неперевірені повідомлення» все одно може бути придатним до дій, якщо його огляд призводить до одного із зафіксованих результатів, наприклад до виправлення раніше поширеного твердження; неперевірені чутки, огляд яких не призводить до жодного з них, придатними до дій не вважаються.

6 У чому аналітики розійшлися

Розбіжності розподілилися нерівномірно. Більшість із них зосередилася на трьох видах граничних випадків:

- Чутки, які згодом можуть підтвердитися: повідомлення непідтверджене на момент перегляду, але досить правдоподібне, щоб згодом знайти підтвердження. Один аналітик може вважати, що на нього варто зреагувати на випадок, якщо воно правдиве; інший може відкласти його до підтвердження.
- Застарілі матеріали, коли інформація точна, але вже неактуальна. Аналітики можуть по-різному оцінювати, наскільки давнім має бути

повідомлення, щоб діяти на його основі вже не мало сенсу.

- Повідомлення поза тематикою, але дотичні до неї: вони виходять трохи за межі завдань служби, водночас пов’язані з ними. Чи вважати таке повідомлення придатним до дій, залежить від того, наскільки вузько кожен аналітик тлумачить ці межі.

Спільне в цих випадках те, що саме повідомлення залишає простір для обґрунтованого судження. Письмові настанови для таких випадків і визначена особа, яка ухвалює щодо них рішення, розв’язують цю проблему безпосередньо.

7 Наслідки для тріажу

Жодна модель тріажу не може бути надійнішою за свої позначки. Модель, навчена й перевірена на позначках, щодо яких двоє аналітиків погоджуються лише частково, зустріне ті самі розбіжності у власних помилках, і частина того, що виглядає як помилка моделі, насправді є невизначеністю самих позначок. Тому заявлені показники моделі слід читати разом із показником узгодженості.

> **Примітка до цього видання (листопад 2026 р.)**
>
> Цю стелю зазначено як обмеження в методологічній записці «Прізми» *«Тріаж для перевантаженої черги»* (2026).

Оскільки розбіжності зосереджуються в граничних випадках, які можна визначити, єдина точка ухвалення рішень щодо них робить робочі позначки послідовнішими. Цей висновок зумовив роль координатора служби в розв’язанні розбіжностей щодо граничних матеріалів.

8 Обмеження

- Дослідження охоплює 200 повідомлень і двох експертів. Один показник к для двох експертів — корисний орієнтир, а не точна чи загальна міра.
- Узгодженість різнитиметься залежно від регіону й категорії, а також із часом, у міру того як змінюється війна. Показник описує цю вибірку й цей період; не варто вважати, що він справджуватиметься деінде чи пізніше.
- Позначки були бінарними: «придатне до дій» або «непридатне». Дослідження не вимірює, наскільки впевнено аналітики дотримувалися своїх суджень чи як би вони ранжували повідомлення.
- Показники за категоріями (таблиця 2) розраховують на частинах вибірки, тож вони менш точні, ніж загальний к.

9 Терміни

Придатне до дій	Повідомлення, перегляд якого має привести до подальших дій служби. Робоче визначення, яке використовували експерти, див. у розділі 2.
Міжекспертна узгодженість	Міра того, наскільки однаково різні експерти оцінюють ті самі матеріали, працюючи незалежно.
Проста узгодженість	Частка матеріалів, яким експерти поставили однакову позначку, без поправки на випадковість.
к Коена	Показник узгодженості двох експертів із поправкою на збіги, очікувані випадково.
Базовий рівень	Частка матеріалів, що належать до певного класу, тут — частка повідомлень, придатних до дій.
Стратифікована вибірка	Вибірка, сформована окремо з кожної групи (тут — за категорією повідомлень), щоб усі групи були представлені.
Розв’язання розбіжностей	Ухвалення остаточного рішення щодо позначки, коли експерти не погодилися; у цьому дослідженні — координатором служби.

Дослідницький центр «Прізма» · Україна · research@prismresearch.org.ua · prismresearch.org.ua