

KSE

Kyiv
School of
Economics



Center for
Sociological
Research

АНАЛІЗ МЕДІЙНОГО ВИСВІТЛЕННЯ СОЦІОЛОГІЧНИХ ОПИТУВАНЬ:

ЛЮТИЙ 2025 РОКУ

Дмитро Андреев

Валентин Гацко

ВСТУП

Опитування громадської думки формують основу для політичних рішень, їм довіряють виборці та науковці, ними апелюють під час дискусій. Оскільки більшість людей знайомляться з результатами опитувань через новини, важливо зрозуміти, як медіа транслюють цю інформацію.

У цьому звіті ми аналізуємо понад 4 500 онлайн-повідомлень лютого 2025 року, щоб показати, які опитування фігурували в українських ЗМІ, які теми найбільше висвітлювались та наскільки публікації дотримуються стандартів методологічної прозорості.

Ми вважаємо, що цей аналіз, окрім задоволення дослідницької допитливості, є важливим з таких причин:

1. Надійні опитування громадської думки є фундаментом демократичного процесу.

«Якщо демократія ґрунтується на волі народу, хіба ми не повинні з'ясувати, у чому ця воля полягає?» — запитував Джордж Ґеллап, засновник опитувань громадської думки. І справді: опитування є інструментом фіксації суспільних настроїв і потреб, на які спираються політики під час розробки програм, законодавчих ініціатив та передвиборних кампаній. Коли результати репрезентативні та методологічно прозорі, влада отримує об'єктивний «зворотний зв'язок», а виборці — уявлення про настрої своїх співгромадян і підстави для свідомого вибору. Навпаки, відсутність надійних даних створює поле для спекуляцій і підриває довіру до інститутів.

2. Медіа формують суспільний порядок денний через вибіркове висвітлення тем опитувань.

Новинні видання і онлайн-портали вирішують, які дослідження опиняться в центрі уваги, а які залишаться в тіні. Якщо акцент ставиться переважно на рейтингах довіри до політичних лідерів чи зовнішню політику, критично важливі питання освіти, охорони здоров'я, екології чи технологічного розвитку опиняються поза фокусом. Така селективність стримує якісне обговорення довгострокових реформ і не дозволяє суспільству сформувати повну картину власних пріоритетів.

3. Методологічна прозорість захищає від маніпуляцій.

Деталі про вибірку, терміни збору даних, похибку та спосіб опитування дають змогу читачам і експертам самостійно оцінити достовірність результатів. У контексті війни та активного поширення дезінформації відкритість методології є запорукою проти маніпуляцій даними. Без базової методологічної інформації складно відрізнити фахове дослідження від маніпулятивної публікації.

Кому може бути корисний цей звіт

- **Журналістам та редакціям** — для вдосконалення стандартів подання результатів опитувань, впровадження перевірених практик і підвищення довіри читачів.
- **Соціологам і дослідникам** — для розуміння тих тем і форматів, які потрапляють у медіа, й ідентифікації «темних плям» у дослідженнях.
- **Політикам і державним органам** — щоб орієнтуватися на актуальні суспільні запити та коректно комунікувати результати опитувань у власних меседжах.
- **Громадським активістам і аналітикам** — для виявлення тем, які потребують додаткового висвітлення, і побудови більш обґрунтованої аргументації у захисті та просуванні соціальних ініціатив.
- **Освітянам та медіаграмотним ініціативам** — як навчальний кейс для прикладу інтерпретації даних і розвитку критичного підходу до споживання інформації.

КОРОТКІ ВИСНОВКИ

- У лютому 2025 року медійний ландшафт соціології в Україні був майже синонімом KMIC: кожна сьома згадка про опитування стосувалася саме його (15,2% усіх випадків). Разом із «Рейтингом» (4,3%) і SOCIS (2,3%) ці три установи утворюють умовне «ядро» української соціології в медіа.
- Серед міжнародних гравців найчастіше згадували YouGov, а також Gallup, Pew Research та Opinion24 — переважно у контексті підтримки України за кордоном. • У публікаціях регулярно з'являлися державні установи — Мінфін (73 згадки, 2,1 %), МОН (26), НБУ (17) та Український ветеранський фонд (22). Це свідчить, що органи влади дедалі частіше замовляють або комунікують результати опитувань.
- Медіа фокусувалися насамперед на політиці та міжнародці:
 - 28% згадок стосувалися довіри до лідерів — Зеленського, Буданова, Залужного — і дискусій про вибори під час війни.
 - 16% — про підтримку України за кордоном, ставлення до переговорів і союзників.
 - 11% — про соціальні проблеми: тривожність, булінг, навантаження на соціпрацівників.
- Медіа значно рідше висвітлюють опитування з тем екології, охорони здоров'я, освіти та технологій. Це може вказувати на низький інтерес журналістів або на відсутність замовників у цих сферах. Водночас у цих напрямках залишається невикористаний потенціал для досліджень, що виходять за межі типових сюжетів, заповнюють прогалини в даних і сприяють формуванню ефективніших політик.
- Найбільший резонанс мали такі сюжети:
 - Булінг серед дітей — 59% школярів стикались із цькуванням, за даними Rakuten Viber.
 - Психічне здоров'я підлітків — масштабне опитування 11–17-річних.
 - Підтримка України у США та Польщі — аргументи у відповідь на заяви Трампа.
 - Вибори під час війни — більшість українців проти.
 - Атомна енергетика — 80% підтримують добудову ХАЕС.
- Журналісти охоче цитують цифри — у 78% новин згадується результати опитувань у відсотках, у 72% — назва організації, що проводила або замовила опитування. Інші деталі методології згадуються рідше:
 - Дати опитування — згадуються лише у 42% випадків.
 - Розмір вибірки — 27%, замовник опитування — 28%.
 - Метод збору даних, похибка вибірки, зважування даних — згадуються надзвичайно рідко (менше ніж у 10% текстів).

МЕТА І МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Метою цього дослідження є виявлення особливостей висвітлення соціологічних опитувань у медійному полі України. Для досягнення цієї мети були сформульовані такі основні дослідницькі завдання:

1. Визначити ключових замовників і постачальників соціологічних даних, присутніх у медійному просторі.
2. Проаналізувати тематичну спрямованість соціологічних опитувань, які висвітлюються українськими ЗМІ.
3. Виявити та охарактеризувати найбільш резонансні випадки використання результатів соціологічних досліджень у медіа.
4. Оцінити повноту та якість висвітлення методологічної інформації щодо соціологічних опитувань.

Для аналізу використовували масив із **4 535 новинних повідомлень**, опублікованих у **лютому 2025 року** на українських медіаресурсах. До аналізу включено лише **новинні матеріали** з таких типів джерел:

- **Інтернет-ЗМІ** — 4 275 повідомлень.
- **Інформаційні агентства** — 236 повідомлень.
- **Друкована преса** — 24 повідомлення.

Публікації були зібрані за допомогою медіамоніторингу за ключовими словами, що вказують на наявність соціологічної інформації: *опитування, соціологічне дослідження і дослідження громадської думки*. До аналізу включалися лише ті повідомлення, де соціологічне опитування або його результати були центральною темою.

Методи обробки поєднували автоматизований аналіз текстів за допомогою моделей штучного інтелекту (GPT-4o/mini) з ручною верифікацією результатів. Зокрема, ми:

- Відібрали релевантні новини, в яких результати опитувань відігравали ключову роль;
- Виокремили назви організацій, що проводили або замовляли опитування, а також короткий зміст теми кожного опитування;
- Стандартизували і уніфікували назви організацій;
- Класифікували тематику кожного дослідження за закритим переліком із 17 тем.
- Оцінили наявність 11 ключових елементів методологічної інформації, серед яких: замовник опитування, виконавець, дати проведення, опис генеральної сукупності, розмір і метод вибірки, похибка, застосування зважування, спосіб проведення, формулювання запитання та представлення результатів опитування.

Детальний опис процедур, промптів і алгоритмів обробки подано в **Додатку**.

РЕЗУЛЬТАТИ

1. Частота згадок окремих організацій у лютому 2025 року.

Новинні видання і онлайн-портали вирішують, які дослідження опиняться в центрі уваги, а які залишаться в тіні. Якщо акцент ставиться переважно на рейтингах довіри до політичних лідерів чи зовнішню політику, критично важливі питання освіти, охорони здоров'я, екології чи технологічного розвитку опиняються поза фокусом. Така селективність стримує якісне обговорення довгострокових реформ і не дозволяє суспільству сформулювати повну картину власних пріоритетів.

У лютому 2025 року в українських медіа **найчастіше** згадували Київський міжнародний інститут соціології (KMIC) — **15.2%** (535 згадок). Це свідчить про його провідну роль як джерела експертної інформації та найбільш впізнаваної дослідницької установи у публічному просторі.

Другу позицію посіла соціологічна група “Рейтинг” — 4.3% (152 згадки), **третю** — міжнародна компанія YouGov (2.5%, 87), а далі йдуть український центр SOCIS (2.3%, 80) і Міністерство фінансів України (2.1%, 73). Сукупно KMIC, Рейтинг і SOCIS формують ядро українських постачальників соціологічних даних, які найчастіше згадувалися в медіа.

Інші **українські дослідницькі та аналітичні центри** також були помітні у публічному обговоренні:

- Центр Разумкова — 0.8% (29)
- Info Sapiens — 0.7% (26)
- Фонд “Демократичні ініціативи” — 0.7% (24)
- Rating Lab — 0.7% (24), нова дослідницька лабораторія, пов'язана з групою “Рейтинг”
- Інститут масової інформації (IMI) — 0.5% (17), що спеціалізується на дослідженнях у сфері медіа.

Серед **міжнародних дослідницьких організацій** в українських медіа фігурували:

- YouGov¹ — 2.5% (87)
- Gallup², Pew Research Center³, Forsa Institute⁴ та Opinia24⁵ — по 0.6% згадувань.

Ці установи є професійними дослідницькими центрами, що регулярно публікують репрезентативні опитування в різних країнах. Згадки стосувалися переважно тем підтримки України за кордоном, міжнародних оцінок війни, політичного лідерства та громадських настроїв.

¹ Міжнародна дослідницька та аналітична група (Велика Британія).

² Американська аналітична та консалтингова компанія, відома глобальними опитуваннями громадської думки (США).

³ Безпартійний американський think tank, що досліджує суспільні проблеми, громадську думку та демографічні тенденції (США).

⁴ Німецька компанія з дослідження ринку та громадської думки (Німеччина).

⁵ Польське агентство з дослідження ринку та громадської думки (Польща).

Окремо згадувалась агенція новин Reuters (1.5%, 52 згадки), яка не є дослідницькою інституцією, але поширює результати опитувань як частину своєї аналітики.

Європейська Бізнес Асоціація (ЄБА) — 0.8% (27 згадок) — фігурувала як замовник досліджень у сфері бізнес-клімату. Хоча вона не є дослідницьким інститутом у вузькому сенсі, її роль як стейкхолдера публічних опитувань є помітною.

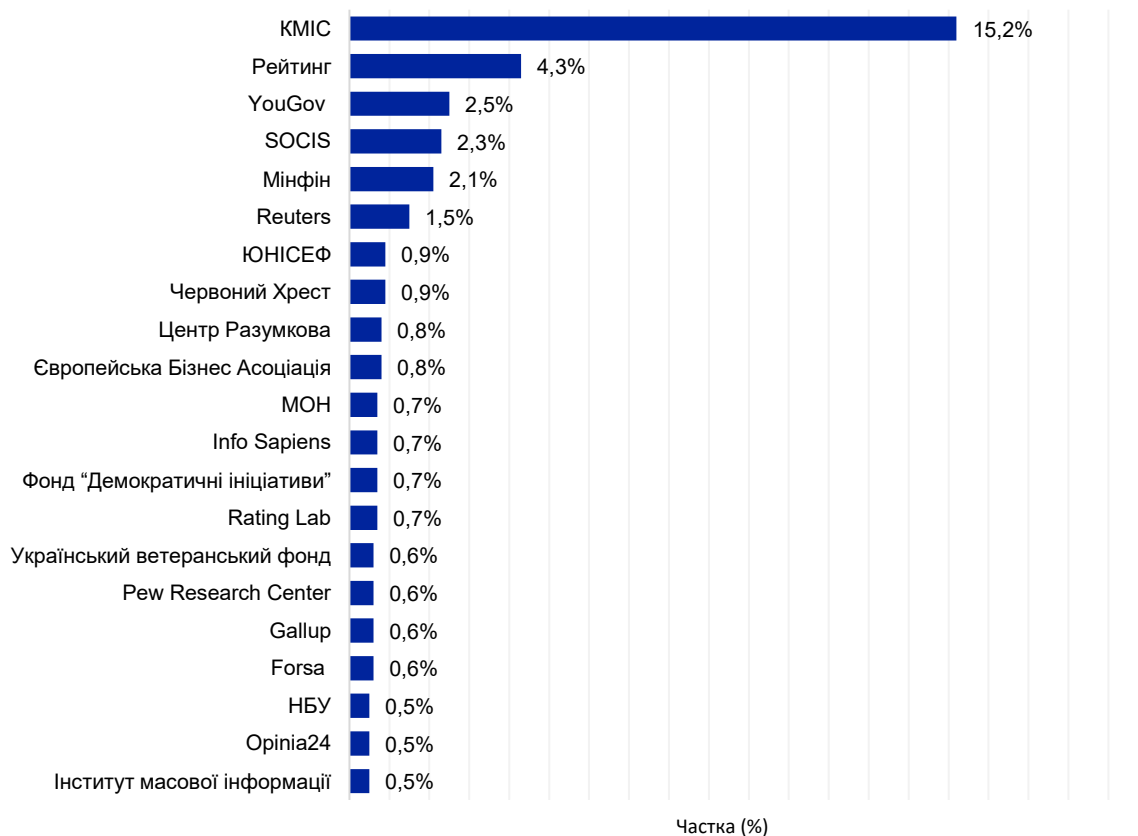
Також помітною є присутність **міжнародних гуманітарних організацій**, таких як Червоний Хрест і ЮНІСЕФ - по 0,9% (31 згадка). Згадки цих організацій пов'язані з опитуваннями у сфері гуманітарного реагування, потреб внутрішньо переміщених осіб та захисту прав дітей.

Серед **установ публічного сектору**, окрім вже згаданого Міністерства фінансів, часто згадувалися:

- Міністерство освіти і науки України — 0.7% (26)
- Національний банк України — 0.5% (17)
- Український ветеранський фонд — 0.6% (22), що також виступав замовником або партнером досліджень.

Це свідчить про зростаючу залученість державних органів до замовлення або комунікації результатів соціологічних опитувань. Це може означати, що влада дедалі активніше використовує соціологію, щоб пояснити свої рішення, підтримати політику або краще зрозуміти ситуацію в країні.

Частка згадок виконавців та замовників досліджень у лютому 2025 року.



2. Про які опитування пишуть медіа?

У лютому 2025 року **найбільше медійних згадок про соціологічні опитування стосувалося тем політичної довіри, міжнародних відносин і соціального добробуту**. Ці три напрямки охопили сукупно понад 55% згадок. Водночас низка важливих суспільних напрямів — освіта, охорона здоров'я, екологія, технології, інфраструктура — залишалися малопредставленими в публічному обговоренні.

Найбільша частка припадає на тему **політичної довіри та врядування** — 28% усіх згадок. У публікаціях обговорювали довіру до президента Володимира Зеленського, голови ГУР Кирила Буданова та головнокомандувача Валерія Залужного. Згадувалися дані КМІС щодо рівня підтримки президента, результати групи «Рейтинг» про 65% довіри до Зеленського, опитування Socis щодо сприйняття легітимності влади в умовах воєнного стану, а також дослідження IRI про ставлення до демократії та переговорного процесу.

На другому місці — **міжнародні відносини** (16%). Ця тема охоплює оцінку союзників, ставлення до міжнародної допомоги чи переговорів з Росією. У новинах, зокрема, цитували результати Gallup щодо ставлення американців до України. Наводилися також дані польського опитування Opinia24 про те, що більшість поляків очікують швидкого мирного врегулювання війни. Крім того, згадувалося масштабне дослідження Фонду Еберта в 14 країнах про геополітичні орієнтації, безпекові страхи й сприйняття Росії як загрози.

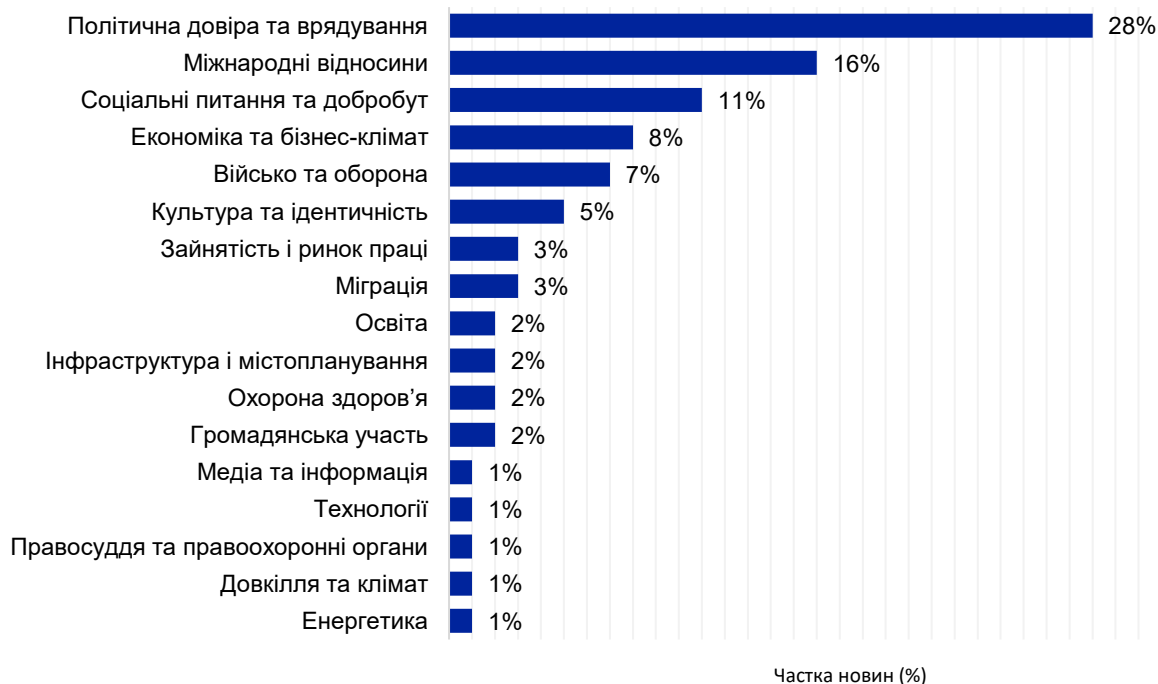
Соціальні питання та добробут становили 11% згадок. Тематика соціального добробуту охоплює питання стресу, тривожності, якості соціальних послуг і підтримки вразливих груп. Згадувалося дослідження Cedos про навантаження на соцпрацівників у Миколаєві, опитування КМІС і Ruban Litvinova про рівень щастя й тривожності серед українців, а також wellbeing-скринінг Betobee щодо добробуту працівників у корпоративному секторі.

Економіка та бізнес-клімат згадувалися в 8% публікацій. До цієї теми належать, наприклад, опитування ЄБА про мобілізаційні виклики для компаній, вплив воєнних ризиків на бізнес-плани та очікування підприємців, а також загальнонаціональне опитування «Рейтингу» про головні загрози для громадян, серед яких на першому місці — економічна криза і зростання цін.

Тематика **військових і оборонних питань** охопила 7% згадок. Вона включала опитування про підтримку української армії, ставлення до мобілізації, бойового духу та довіру до військових лідерів.

Інші тематичні напрями згадувались рідше. Зокрема, культура та ідентичність охопила близько 5% згадок, зайнятість, ринок праці та міграція — близько 3%. Решта тем, включно з освітою, інфраструктурою, охороною здоров'я, громадянською участю, медіа та інформацією, енергетикою, технологіями, довкіллям і правосуддям, згадувались менш ніж у 3% повідомлень.

Найпопулярніші теми опитувань у лютому 2025.



3. Теми найбільш “гучних” повідомлень про опитування

ТОП-теми лютого за охопленням (за змістом 10 найпопулярніших повідомлень):

№	Тема	Короткий зміст
1	Булінг серед дітей	Результати опитування Rakuten Viber, згідно з яким 59% дітей стикалися з булінгом, подано в новині як вступ до заяв омбудсмена про недостатню ефективність захисту дітей від цькування.
2	Психічне здоров'я підлітків	Всеукраїнське опитування підлітків 11–17 років висвітлює дані щодо психічного здоров'я, булінгу, ризикованої поведінки та впливу війни. Результати подають актуальний портрет стану молоді в Україні після пандемії COVID-19 та в умовах повномасштабної війни.
3	Міжнародна підтримка України	Опитування про підтримку України з боку громадян США згадується як аргумент у відповідь на заяви Дональда Трампа, підкреслюючи, що більшість американців залишаються на боці України.
4	Ставлення до атомної енергетики	Опитування групи "Рейтинг", за яким 80% українців підтримують добудову ХАЕС, а 69% оцінюють енергетичну політику влади як ефективну, використовується як аргумент на підтримку рішень уряду в енергетичній сфері.
5	Рівень довіри до президента та підтримка ідеї виборів під час війни	Результати опитувань SOCIS і KMIC, які демонструють високий рівень довіри до президента Зеленського, спростовують заяву Трампа про 4% підтримки. Дані також вказують на небажання українців проводити вибори до завершення війни.

№	Тема	Короткий зміст
6	Вибори під час війни	Опитування SOCIS показує, що понад 63% українців виступають проти проведення виборів до завершення війни. Ці результати активно згадуються у відповідь на заяви Трампа та в контексті дискусій про зовнішній тиск щодо виборів в Україні.
7	Освітні потреби учнів 5–11 класів у межах реформи "Освіта для життя".	Міністерство освіти і науки у партнерстві з U-Report проводить опитування учнів 5–11 класів для з'ясування їхніх освітніх потреб. Його результати мають стати основою для реалізації політики "Освіта для життя".
8	Ставлення українців до політичних лідерів і форматів мирних переговорів.	Опитування групи "Рейтинг" демонструє позитивне ставлення українців до низки європейських лідерів і високий рівень довіри до президента Зеленського (65%). Більшість респондентів підтримують переговори України з РФ, але виступають проти перемовин США з Росією без участі України.
9	Ставлення громадян Польщі до можливої стратегії України у війні з Росією.	Опитування Opinia24 показує, що 57% поляків підтримують швидке укладення миру Україною, тоді як 30% виступають за продовження бойових дій. Погляди респондентів розрізняються залежно від статі та політичних уподобань. Результати наведені в новині для ілюстрації змін у громадській думці в Польщі.
10	Ставлення громадян Німеччини до майбутньої підтримки України урядом ФРН.	Згідно з опитуванням Insa, 46% німців не підтримують подальше надання зброї чи фінансової допомоги Україні, тоді як 28% виступають за збереження підтримки. Погляди суттєво відрізняються за віком, статтю, регіоном та партійною орієнтацією. Дані наведено як приклад змін у ставленні громадськості в Німеччині напередодні формування нового уряду.

4. Якість публікацій про соціологічні опитування

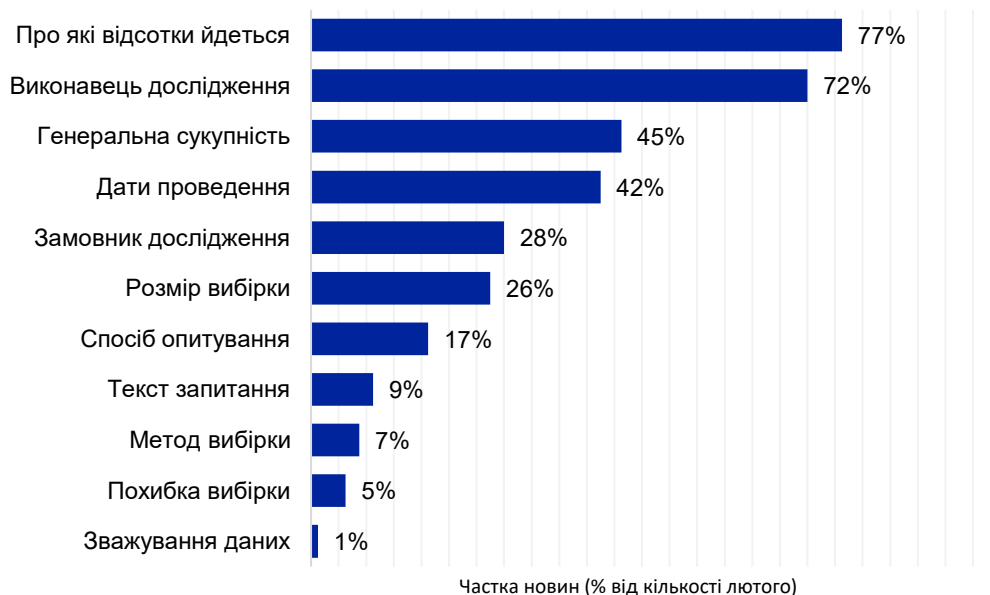
Українські медіа зазвичай повідомляють лише результати опитувань, але не пояснюють, як ці опитування проводились. Така ситуація не унікальна для України, але суттєво впливає на довіру до соціологічної інформації в публічному просторі.

Аналіз медійних згадок показує, що лише окремі методологічні деталі соціологічних опитувань стабільно фігурують у публікаціях. Найчастіше вказуються **відсотки, на яких ґрунтуються висновки** (78%) та **організація, яка проводила опитування** (72%). Ці дві позиції формують основу більшості новинних повідомлень про опитування.

Менш ніж у половині випадків згадуються **дати проведення опитування** (42%) та **генеральна сукупність** (46%), що вже суттєво обмежує можливість оцінити репрезентативність. Ще рідше трапляються згадки про **розмір вибірки** (27%) та **замовника дослідження** (28%), хоча саме ці характеристики часто викликають запитання щодо достовірності результатів.

Найменше уваги приділяється **способу проведення опитування** (17%), **тексту запитання** (9%), **методу вибірки** (8%), **похибці вибірки** (6%) та **використанню вагових коефіцієнтів** (1%). Відсутність цих критично важливих елементів значно ускладнює оцінку якості даних і може сприяти поширенню некоректних інтерпретацій результатів.

Частка новин у лютому, що згадують методологічну інформацію.



Порівняння з міжнародним контекстом

Порівняння з іншими країнами слід тлумачити з обережністю, оскільки українські дані ґрунтуються переважно на онлайн-публікаціях за один місяць (лютий 2025 року). Натомість більшість міжнародних досліджень охоплюють інші типи медіа (друковану пресу, телебачення) і часто фокусуються на виборчих кампаніях. Ці фактори суттєво впливають на частоту, глибину та характер подачі соціологічної інформації, а отже, пряме порівняння може бути некоректним. Проте такі зіставлення дозволяють окреслити загальні тренди прозорості у висвітленні результатів опитувань.

Онлайн-медіа в Україні рідше повідомляють ключові методологічні деталі соціологічних опитувань, ніж великі національні онлайн-медіа США під час виборів. Найближчими за логікою вибірки є дослідження Turcotte et al. (2017) та Kim et al. (2010), які кодували новини у великих національних онлайн-медіа США у період виборчих кампаній (праймериз 2016 року та президентська кампанія 2008 року відповідно). У цих дослідженнях рівень методологічної прозорості був суттєво вищим, ніж в українських онлайн-медіа. Наприклад, замовник опитування згадувався у майже 100% новин у США, тоді як в Україні — лише у 28% випадків. Дати проведення, розмір вибірки та похибка вибірки згадувалися в українських публікаціях у 42%, 27% і 6% випадків відповідно, тоді як у США — в діапазоні 54–92%. Єдине поле, де українські медіа не поступаються, — це згадка про формулювання запитання (9% в Україні проти 5–8% у США). Водночас українські видання частіше зазначають, яка саме організація проводила опитування (72%), однак у згаданих американських дослідженнях цей параметр не вимірювався, що унеможливлює пряме порівняння.

Міжнародні дослідження показують, що тип медіа і виборчий контекст суттєво впливають на рівень прозорості, що частково пояснює нижчі показники в українській вибірці. Зокрема, друковані ЗМІ зазвичай демонструють вищі показники методологічного

розкриття, ніж онлайн-видання (Larsen & Straubinger, 2012). Крім того, в періоди виборчих кампаній медіа приділяють більше уваги достовірності соціологічних даних, що підвищує ймовірність згадок про дати опитування, похибку, вибірку тощо (Miller & Hurd, 1982; Gosselin & Pétry, 2009). Натомість українська вибірка охоплює лише онлайн-видання і звичайний, невиборчий період — лютий 2025 року. Отже, **існує висока ймовірність, що розрив у прозорості є частково результатом відмінностей у контексті досліджень, а не лише проблемою українських медіа**. Якби порівняння здійснювалося з не-виборчими новинами в онлайн-медіа інших країн, результати могли б виявитися ближчими.

Порівняння розкриття методологічної інформації в медійному висвітленні опитувань у різних дослідженнях.

Джерело дослідження	Країна	Організація, що провела	Замовник (спонсор)	Дати збору даних	Розмір вибірки	Генеральна сукупність	Формулювання запитання	Метод збору	Похибка	Метод вибірки	Вагування
Україна (2025)	Україна	72	28	42	27	46	9	17	6	8	1
Kim (2011)	США	-	100	84	84	84	8	-	92	77	-
Turcotte et al (2017)	США	-	99	56	57	83	5	22	54	11	-
Vögele & Bachl (2018)	Німеччина	84	-	31	26	42	55	6	4	-	-
Brettschneider (1997)	Німеччина	93	57	63	-	72	33	52	3	-	-
Ferguson & de Clercy (2005)	Канада	91	60	46	0	20	20	-	52	-	-
Andersen (2000)	Канада	95	-	64	33	-	21	-	25	-	-
Hardmeier (1999)	Швейцарія	81	47	44	51	-	30	11	4	4	32
Strömbäck (2009)	Швеція	-	85	-	-	35	40	26	-	-	-
Weaver & Kim (2002)	США	-	78	39	37	30	6	21	12	-	-
Welch (2002)	США	-	89	43	38	54	10	10	38	6	2
Portilla (2016)	Іспанія	83	85	77	85	56	60	52	46	38	27
De Vreese & Semetko (2002)	Данія	-	95	87	92	3	87	84	84	-	-
Cheng (1999)	Сінгапур	-	-	51	-	98	-	17	0	34	-
Sonck & Loosve ldt (2008)	Бельгія	16	18	9	33	7	4	12	-	2	-
Szwed (2011)	Польща	92	59	-	35	45	13	7	6	54	-

ДОДАТКИ

Технічні деталі аналізу

1. Вибір релевантних повідомлень

На першому етапі модель GPT аналізувала повний текст кожного повідомлення, щоб визначити, чи є соціологічне опитування або його результати **центральною темою** повідомлення. Використовувався короткий промпт із двома можливими відповідями ("так" або "ні"). Повідомлення, в яких опитування не було основним фокусом, відфільтровувалися.

Модель GPT: gpt-4o-mini.

Промпт:

*Чи є соціологічне опитування або його результати ****центральною темою**** цього повідомлення?*

Відповідай лише 'Так' або 'Ні'.

Текст повідомлення:

[txt]

2. Виділення організації, що проводила або замовляла опитування, та короткого опису теми опитування.

Для повідомлень, що пройшли попередню фільтрацію, використовувався один комбінований промпт для одночасного вилучення назви організації, що проводила або замовляла опитування, та короткого опису теми опитування. Хоча у формулюванні промпта ми запитуємо організацію, яка проводила опитування, модель на практиці виділяє як виконавця, так і замовника дослідження.

Модель GPT: gpt-4o-mini.

Промпт:

1) *Вкажи виконавця опитування у форматі `Хто проводить: <назва>`.*

2) *Одним реченням дай коротку тему опитування у форматі `Тема: <тема>`. Відповідай рівно двома рядками, без жодних інших пояснень.*

[text]

3. Валідація, очищення та нормалізація назв організацій

Спочатку GPT перевіряв, чи є витягнута назва конкретною назвою організації. Далі застосовувались текстові правила для видалення зайвих слів, символів, регістрів тощо. Для деяких випадків використовувались вручну задані винятки — список назв, які необхідно зберігати, попри класифікацію моделі.

Промпт:

Чи наведена тут конкретна назва установи, організації чи видання?

Якщо так, то поверни назву, якщо ні, то напиши 'ні' одним словом.

Output format: Назва установи/організації/видання або 'ні'.

[txt]

4. Групування та уніфікація назв організацій.

Цей етап передбачав кілька послідовних кроків:

4.1: Початкове групування на основі входження рядків.

Унікальні назви організацій (окрім "ні" та стоп-слів типу "центр", "інститут") групувалися, якщо одна назва містилася в іншій. Групи сортувалися за довжиною назв всередині групи та за першим елементом між групами.

4.2: Ітеративне попарне порівняння та уніфікація за допомогою GPT:

Відсортований список згрупованих назв проходив ітеративну обробку. Пари сусідніх назв подавалися GPT для визначення, чи позначають вони ту саму організацію, і якщо так, то яка з назв є коротшою/коректнішою.

Модель GPT: gpt-4o.

Промпт:

Порівняйте наступні дві назви організацій:

1) '[prev_name]'

2) '[next_name]'

Якщо ці назви позначають ту ж саму організацію, і одна з них є коротшою, поверніть коротшу назву. Якщо це різні організації, поверніть першу назву.

4.3: Фінальне групування топ-150 назв за допомогою GPT

Було сформовано список 150 найчастіше згадуваних назв організації. Список топ-150 назв подавався GPT з проханням упорядкувати його так, щоб дублікати йшли один за одним.

Модель GPT: gpt-4o.

Промпт:

Ось список 150 назв організацій, впорядкованих за частотою:

[top_150_names]

Будь ласка, проаналізуйте цей список і, якщо знайдете дублікати (назви, що позначають ту ж організацію), упорядкуйте їх так, щоб повторювані назви йшли один за одним.

Вихідний формат: список назв (без нумерації), по одному рядку.

4.4: Кластеризація та вибір представницької назви GPT.

Очищений та впорядкований список назв подавався GPT для кластеризації та вибору найкоротшої представницької назви для кожного кластера.

Модель GPT: gpt-4o.

Промпт:

Ось список назв організацій:

[good_lines]

Групуйте всі варіанти назв однієї організації в кластери. Для кожного кластера оберіть найкоротшу назву як представницьку і виведіть рядок у форматі:

представницька (варіант1, варіант2, ...)

Якщо кластер містить лише одну назву, виведіть її без дужок.

Відповідь GPT парсилася для створення таблиці відповідності "варіант назви -> канонічна назва". Ця таблиця використовувалася для оновлення назв організацій.

5. Аналіз якості подання результатів опитувань.

Для кожного повідомлення ми визначали наявність 11 типів методологічної інформації (замовник, виконавець, дати, генеральна сукупність, розмір вибірки, метод вибірки, похибка, зважування, спосіб опитування, текст запитання, згадка відсотків).

Модель GPT: gpt-4o-mini.

Промпт:

Для кожного з наведених пунктів напиши 'так' або 'ні' — чи згадується ця інформація в повідомленні. Поверни рівно 11 рядків у форматі:

Замовник опитування: так/ні

Організація, яка проводила опитування: так/ні

Дати проведення опитування: так/ні

Генеральна сукупність: так/ні

Розмір вибірки: так/ні

Метод вибірки: так/ні

Похибка вибірки: так/ні

Застосування вагових коефіцієнтів: так/ні

Спосіб проведення опитування: так/ні

Текст запитання: так/ні

Про які відсотки йдеться: так/ні

Текст повідомлення:

[text]

6. Класифікація теми дослідження (за закритим переліком):

Тематика кожного опитування була класифікована за одним із 17 попередньо визначених напрямів з дефініціями (див. *Таблиця 1. Тематика опитувань та їх визначення*). На вхід моделі для класифікації подавався короткий опис теми, вилучений на етапі 2, а не повний текст повідомлення.

Модель GPT: gpt-4o-mini.

Промпт:

Classify the survey's primary topic from these categories:

[список_тем_через_крапку_з_комою] .

Definitions:

[список_визначень_тем_через_крапку_з_комою] .

Survey description:

[text_from_survey_topic_column]

Return only the best-matching topic or 'None'.

Таблиця 1. Тематика опитувань та їх визначення

Категорія	Що охоплює
Довіра до політиків та інститутів, і врядування	Довіра до органів влади, політичних лідерів, державних інституцій, політичні настрої, врядування.
Зовнішні відносини / Міжнародні справи	Міжнародна дипломатія, зовнішня політика, відносини між країнами, геополітика, міжнародні організації.
Військо та оборона	Збройні сили, національна безпека, бойові дії, військова політика, ставлення до армії.
Економічний та бізнес-клімат	Економічні тенденції, бізнес-середовище, інвестиції, підприємництво, ринок.
Зайнятість і ринок праці	Рівень зайнятості, ринок праці, безробіття, умови праці, трудові права.
Соціальні проблеми та добробут	Бідність, нерівність, соціальний захист, демографія, вразливі групи, якість життя.
Охорона здоров'я	Медичні послуги, здоров'я населення, інфраструктура, страхування, захворювання.
Освіта	Школи, університети, якість освіти, грамотність, політика в освіті, інфраструктура.
Інфраструктура та міське планування	Транспорт, дороги, житло, розвиток міст, комунальні послуги, просторове планування.
Громадянська участь	Участь у виборах, громадська активність, волонтерство, взаємодія з владою.
Культура та ідентичність	Національна ідентичність, традиції, мова, історична пам'ять, культурні практики.
Правопорядок і судочинство	Злочинність, правоохоронна система, суди, безпека громадян, поліція, тюрми.
Довкілля та клімат	Екологія, зміна клімату, забруднення, охорона природи, стале використання ресурсів.
Медіа та	Медіа, довіра до інформації, споживання новин,

Категорія	Що охоплює
інформація	дезінформація, медіаграмотність.
Технології	Технології, цифровізація, інтернет, кібербезпека, інновації, діджитал-інфраструктура.
Енергетика	Політика в енергетиці, джерела енергії, ринки палива, відновлювана енергетика.
Міграція	Міграція, біженці, інтеграція, політика щодо іммігрантів, громадська думка про мігрантів.

Таблиця 2. ТОП10 повідомлень про опитування за місяць

Опис	Сумарна кількість переглядів разом із дублями	Тип ЗМІ	Джерело
ТОП-10 повідомлень за лютий 2024			
Опитування згадується для ілюстрації масштабу проблеми булінгу серед українських дітей, що підкреслює необхідність змін у законодавстві для їх захисту.	525000	Інтернет	vseosvita.ua
Опитування згадується для збору даних про добробут молоді та вплив різних факторів на їхнє життя в Україні.	351058	Інтернет	unicef.org/ukraine
Опитування згадується для підтвердження того, що більшість американців підтримують Україну та вважають Путіна загрозою.	134281	Інтернет	TCH
В цьому повідомленні згадується опитування, щоб продемонструвати високий рівень підтримки українцями добудови атомної електростанції як засобу для стабілізації тарифів на електроенергію.	82653	Інтернет	podrobnosti.ua
Опитування згадується для спростування неправдивих заяв Дональда Трампа про низький рівень підтримки Володимира Зеленського серед українців.	58529	Інтернет	Українська правда.ua
Опитування згадується для ілюстрації соціологічних трендів, які впливають на політичну ситуацію в Україні, зокрема на рейтинг президента Володимира Зеленського та появу нових конкурентів.	57287	Інтернет	Українська правда.ua
Опитування згадується для з'ясування реальних освітніх потреб школярів з метою покращення системи навчання в Україні.	53094	Інформаційне агентство	Interfax-Україна
Опитування згадується для того, щоб продемонструвати ставлення українців до європейських лідерів та політиків, а також їхню думку щодо переговорів з Росією.	45338	Інтернет	bukinfo.com.ua
Опитування згадується, щоб продемонструвати думки поляків щодо поведінки України у війні з Росією.	43163	Інтернет	Європейська правда
Опитування згадується для ілюстрації зміни ставлення німців до підтримки України, що підтверджується даними, отриманими від інституту дослідження громадської думки Insa для t-online.	42152	Інтернет	Європейська правда