



Фінансується
Європейським Союзом

RSF РЕПОРТЕРИ
БЕЗ КОРДОНІВ

ЖІНКИ в медіа.

Шаблон інструкції використання ШІ в редакціях

Цей документ підготовлено за фінансової підтримки Європейського Союзу в межах діяльності «Репортерів без кордонів» в Україні як частина проекту «Не штучні загрози: протидія насильству проти журналісток із використанням ШІ в Україні», що реалізується ГО «Жінки в медіа». Зміст цього документа є виключною відповідальністю ГО «Жінки в медіа» і за жодних обставин не може вважатися таким, що відображає позицію «Репортерів без кордонів» або Європейського Союзу.

Керівництво

1/ Інструкції мають ґрунтуватися на принципах, задекларованих у Політиці використання ШІ. Водночас, на відміну від Політики, вони є прикладним внутрішнім документом – тобто редакції не потрібно їх публікувати чи оформлювати певним чином. Такий документ має бути максимально зручним і практичним, а також адаптованим до потреб, які було визначено перед написанням Політики. Важливо пам'ятати, що:

Політика використання ШІ

→ це загальний рамковий документ, оприлюднений на вебсайті медіа, який окреслює основні цінності, принципи та підходи в цій сфері. Такі політики не є щоденним робочим інструментом, а радше визначають загальний напрям і стандарти.

Інструкція з використання ШІ-інструментів

→ це оперативний документ, який, спираючись на цінності та принципи, визначені в політиці, надає практичні рекомендації щодо використання конкретних ШІ-інструментів для виконання конкретних завдань.

Реєстр ШІ-інструментів

→ перелік інструментів, які використовує медіа із зазначенням мети та масштабів використання.

2/ Для різних інструментів ШІ можуть розроблятися різні інструкції залежно від типу завдань, які виконують за допомогою тієї чи іншої системи. Водночас, існують загальні принципи, яких слід дотримуватися при використанні ШІ. Йдеться про цифрову безпеку, відповідальний вибір систем ШІ та безпечний промпт інжиніринг.

3/ Ця Інструкція надає загальні рекомендації щодо вибору систем ШІ, організації робочого простору в рамках ШІ-сервісів та безпечного дизайну промптів (запитів) відповідно до стандартів прав людини. Важливо, зміст цієї Інструкції є актуальним для:

- працівників, які використовують ШІ-продукти загального користування (на кшталт ChatGPT, Notebook LM, Claude, Perplexity тощо);
- працівників, які використовують системи ШІ, розроблені для конкретного медіа, але при цьому не є технічними спеціалістами (тобто виступають у ролі користувачів, а не розробників чи впроваджувачів конкретної технології).

Якщо медіа планує самостійно розробляти, дотреновувати чи налаштувати системи ШІ, ми радимо звертатися до технічних експертів або мати в штаті людину, яка може здійснювати технічну підтримку.

4/ Інструкції мають створювати залежно від завдань, які медіа ставить перед собою, впроваджуючи ШІ, а також від функціоналу і дизайну конкретного інструменту ШІ.

Шаблон інструкції використання ШІ

Ця Інструкція – прикладний інструмент для вибору і впровадження ШІ, за допомогою якого вирішують конкретні редакційні задачі. В контексті використання ШІ в операційних цілях важливим є коректне налаштування системи, в той час як для генеративного ШІ фокусом має бути навчання команди правильно дизайнувати промпти (запити до системи).

/ 01

При виборі системи ШІ важливо уникати інструментів, які створюють ризики для прав людини, суперечать журналістським стандартам та етиці, або походять з держави-агресора чи держави з високим індексом порушення прав людини. Також при виборі систем ШІ варто брати до уваги:

- кому належать права інтелектуальної власності на створений контент?
- чи мають розробники системи доступ до файлів на вашому пристрої? чи можна обмежити такий доступ?
- чи навчається система на введених даних? чи можна обмежити таку функцію?
- чи виправдано використовувати цей інструмент з огляду на кількість часу, яка витратиться на перевірку результатів його роботи?
- чи є платна версія набагато кращою за безкоштовну? чи може редакція собі дозволити платну версію?
- чи компанія-розробник добросовісна і не має історії скандалів, пов'язаних із порушенням прав людини?

Щоб обрати системи ШІ, які відповідають стандартам у сфері захисту прав людини, радимо користуватися [посібником](#). Важливо, після вибору конкретного інструменту його варто завантажувати лише з офіційних онлайн-магазинів (GooglePlay, AppStore) або відкривати офіційну вебверсію.

Після вибору/запуску інструменту ШІ слід звернутися до технічного експерта для його налаштування або, якщо в команді немає технічного спеціаліста, а медіа використовує систему ШІ загального користування – **зробити загальні кроки для мінімізації ризиків**. Вони передбачають:



Налаштування двофакторної аутентифікації

Для уникнення зламу акаунтів варто вживати заходів з цифрової безпеки. В тому числі, це стосується використання окремих систем ШІ



Заборона можливості навчання системи ШІ на введених даних

Якщо у налаштуваннях системи є можливість вимкнути функцію навчання ШІ на даних користувача – варто нею скористатися.

Особливо це актуально, якщо ви працюєте з чутливими темами. При цьому, це не виключає потреби слідкувати за тим, щоб конфіденційна інформація в цілому не потрапляла в систему ШІ загального доступу;



Налаштування регулярного видалення історії запитів (за потреби)

Деякі інструменти дозволяють регулярно видаляти історію запитів – або зі спливом певного часу (наприклад, 1 місяця), або після закриття вкладення у браузері.

/ 03

Важливо **регулярно проводити навчання команди**. Особливо актуальним це є, якщо інструмент ШІ є складним у використанні, відбулися значні оновлення у функціоналі і попередні тренінги є неактуальними або якщо працівник не має базових навичок роботи з ШІ-інструментами такого типу, що використовуються у редакції. Навчання бажано робити не загальним, а орієнтованим на використання конкретної системи ШІ.

/ 04

Одним з найважливіших аспектів при використанні ШІ є ефективний **промптинг** – вміння поставити для ШІ-інструменту задачу. Перш за все слід **чітко формулювати запит до системи**. Це включає уточнення мети запиту, його контексту та правильного підбору термінів. В цілому, бажано робити запит конкретним, а не загальним.

✘ Неправильно: *Яка відповідальність буває за вирубку лісу?*

✓ Краще: *Я готую аналітичний матеріал для медіа журналістів-розслідувачів на тему корупційних порушень у сфері запобігання вирубці лісів. Знайди мені резонансні іноземні випадки, коли посадовців притягали до відповідальності за надання незаконних дозволів на вирубку лісу.*

Якщо запит є дуже великим за обсягом: наприклад, тема є надміру обширною або включає багато підпунктів – краще розбити таке завдання на декілька дрібніших. Так система видаватиме детальніші і чіткіші результати.

✘ Неправильно: *Яким має бути захист прав дітей в Інтернеті?*

✓ Краще: *Чи зобов'язані соціальні мережі впроваджувати заходи для захисту прав дітей?
Як батьки можуть забезпечити дітей від небезпек в Інтернеті?
Які бувають види заходів для перевірки віку користувачів в мережі?
Які найпоширеніші ризики для дітей у месенджерах?*

Для полегшення перевірки інформації варто додавати у запит вимогу надати джерела інформації, якими користується система ШІ. При цьому, важливо, щоб джерела були у форматі активних посилань на вебресурси, які можна перевірити, перейшовши за ними. Це зекономить час від перехресного пошуку у Google винятково за назвою.

Варто пам'ятати, що ШІ може буквально «сконструювати» травматичний досвід, створити неправдиві деталі, відтворити стереотипні образи, романтизувати або драматизувати насильство або навіть закласти ризик віктимблеймінгу через саму рамку питання.

✘ Неправильно: *Напиши драматичну історію про жертву домашнього насильства*

✓ Краще: *Запропонуй 5 аналітичних ракурсів для статті про домашнє насильство в Україні під час війни, які виходять за межі кримінальної хроніки, уникаючи сенсаціоналізації, віктимблеймінгу та стереотипних описів. Зосередься на економічних бар'єрах для виходу з аб'юзивних стосунків у сільській місцевості та на тому, як переселення змінює динаміку влади в родині. Використовуй підхід, орієнтований на захист прав людини.*

✘ **Неправильно:** Напиши статтю про те, як жінки страждають від насильства під час війни

✔ **Краще:** Склади список із 7 критичних запитань до місцевих органів влади щодо імплементації Стамбульської конвенції в Україні в умовах воєнного стану. Запитання мають стосуватися доступності шелтерів для жінок з інвалідністю та підготовки соціальних працівників у прифронтових регіонах.

ПРОМПТИ ДЛЯ ВІЗУАЛІЗАЦІЇ ТА МЕТАФОР. ШІ часто генерує стереотипні зображення (жінка в синцях, що плаче).

✘ **Неправильно:** Зроби картинку про газлайтинг, щоб вона викликала сильні емоції.

✔ **Краще:** Опиши ідею для ілюстрації на тему психологічного насильства та газлайтингу. Категорично уникай кліше (синяки, кулаки, темні кімнати, закриті обличчя). Запропонуй візуальні метафори, що відображають втрату ідентичності та шлях до відновлення сили потерпілої особи.

ПЕРСОНАЛЬНІ ДАНІ ТА КОНФІДЕНЦІЙНА ІНФОРМАЦІЯ. Для роботи з конфіденційними даними слід використовувати лише локальні системи ШІ – такі, що встановлюються, тренуються та адаптуються до потреб медіа, які повністю займаються адмініструванням таких систем. Якщо працівники використовують системи ШІ загального доступу, жодної інформації з обмеженим доступом в них вноситися не має.

Щодо будь-якої системи ШІ загального користування має існувати презумпція того, що інформація, внесена в цю систему, може стати публічною або загальнодоступною.

При цьому, варто пам'ятати про декілька **важливих правил при формулюванні промптів** (навіть якщо промпти не містять документів з конфіденційною інформацією), які можуть стосуватися чутливих тем:

- не запитувати персональні дані приватних осіб;
- не заохочувати ШІ здійснювати доксинг чи стеження, профайлінг особи тощо;
- під час аналізу реальних випадків використовувати анонімізовані приклади;
- відповідально формулювати гіпотетичні ситуації та запитання.

РІВНІСТЬ, НЕДИСКРИМІНАЦІЯ ТА ЗАПОБІГАННЯ НАСИЛЬСТВУ. При дизайнуванні промптів варто уникати формулювань, які можуть підштовхнути ШІ-інструмент до видачі упереджених результатів. Щоб уникнути таких ситуацій, не варто: фокусуватися на групах крізь призму їх приналежності до раси, релігії, статі, національності, сексуальної орієнтації, наявності інвалідності тощо.

✘ **Неправильно:** Чому ВПО є причиною підвищення рівня злочинності?

✔ **Краще:** Які чинники впливають на підвищення рівня злочинності в регіоні?

Або

✘ **Неправильно:** Чому жінки не йдуть із насильницьких стосунків?

✔ **Краще:** Які системні бар'єри перешкоджають жінкам залишати насильницькі стосунки в умовах війни?

До того ж, варто уникати використання стереотипів та негативного контексту в промпті. Натомість, слід використовувати нейтральну мову та формулювати запити на чутливі теми аналітично, а не ствердно.

✘ **Неправильно:** Чому в Україні нетолерантне ставлення до ЛГБТ-спільноти?

✔ **Краще:** Яка наразі ситуація із захистом прав ЛГБТ-спільноти в Україні? Які проблеми є найбільш вираженими?

Окрім нейтральних та позитивних формулювань, а також деталізації самого запиту, варто також додавати запобіжники від неправильного тлумачення запиту системою ШІ. Зокрема, йдеться про проактивне додавання у промпт таких застережень:

«Уникай стереотипів, упереджень і дискримінаційних висловлювань»;

«Використовуй гендерно чутливу лексику при генеруванні відповіді»;

«Уникай висловлювань та прикладів, які можуть ретравматизувати постраждалих від [...]» (у випадку, якщо йдеться про випадки насильства, постраждалих від злочинів, тощо);

«Уникай деталізованого опису сцен насильства».

АВТОРСЬКЕ ПРАВО. По-перше, варто уважно перечитувати умови користування системою ШІ та використовувати такі інструменти, правовий режим яких підходить для медіа. Наприклад, якщо продукт передбачає, що компанія-розробник залишає за собою виключне право на розпорядження результатами роботи ШІ, це може в подальшому призвести до проблем з монетизацією контенту.

По-друге, щоб система ШІ не видавала результатів, які порушуватимуть авторські права інших осіб, варто дизайнувати промпти відповідально, зокрема, уникаючи прямих відсилок до об'єктів, захищених авторським правом, з метою їх повного або часткового відтворення (якщо таке відтворення прямо не передбачене авторським задумом і не переслідує законну мету).

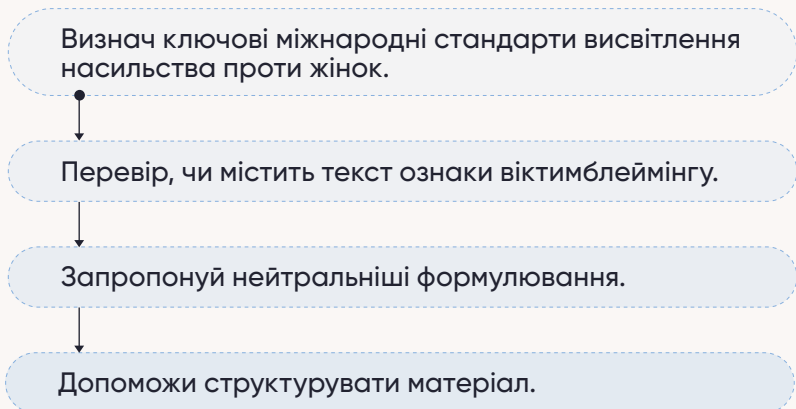
✘ Неправильно: Створи мені статтю для вебсайту медіа про сучасну українську літературу, наслідуючи письменницький стиль Сергія Жадана.

✓ Краще: Згенеруй мені тези/цікаві ідеї для статті на вебсайт медіа про сучасну українську літературу. Використовуй стиль написання, який буде цікавим для онлайн-читачів. Запропонуй цікаві метафори і приклади. Нехай текст не міститиме канцеляриту та бюрократичних термінів.

/ 05

Варто застосовувати принцип **Поетапної взаємодії з генеративним ШІ** (Prompt Laddering) під час роботи з чутливими темами, зокрема насильством проти жінок, сексуальним насильством, війною, дискримінацією та правами людини. Такий підхід передбачає використання серії послідовних, вузько сформульованих запитів замість одного комплексного або неконтрольованого промпту для генерації готового матеріалу.

Наприклад:



/ 06

Зрештою, важливим етапом є перевірка результатів роботи системи ШІ і забезпечення фахового людського нагляду. Якщо йдеться про генеративний ШІ, важливо перевірити:

- **автентичність джерел** (ШІ може галюцинувати і посилатися на неіснуючі ресурси, вигадувати авторів, дослідження чи події);
- **якість джерел, на які посилається ШІ** (навіть коли джерело справжнє, ШІ часто використовує контент, який пріоритезують пошукові системи, на кшталт Google. Такий контент, особливо на політичні та соціальні, може походити з джерел, спеціально створених для поширення дезінформації. Наприклад, так росіяни організували мережу фейкових ресурсів "Правда", які використовують в тому числі, щоб системи ШІ посилалися на неправдивий контент при видачі результатів на основі пошуку в мережі);
- **наявність упреджень у контенті** (навіть при якісних промптах ШІ все ще може послуговуватися дискримінаційною лексикою чи створювати візуальний контент, який підсилює нерівність);
- **коректне використання термінології** (ШІ може послуговуватися синонімами чи зворотами, які є контекстуально недоречними чи неправильними. Особливо актуальною перевірка є, коли йдеться про юридичні тексти).

Також важливо перевіряти, чи ШІ, що використовується в чат-ботах, для розподілу навантаження в команді чи для інших завдань коректно виконує свою роботу. Для цього варто регулярно перевіряти технічні дозволи, надані програмі під час її використання, оновлення умов користування та функціоналу системи ШІ.

