

Ні «підсолоджуванню» зґвалтування

Соціальні мережі змушують замовкнути тих, хто говорить про насилля відверто. Там, де має бути правда, використовуються евфемізми.

Сьогодні життя суспільства значною мірою визначають соціальні мережі. Те, що за початковим задумом було простором розваг і спілкування, перетворилося на поле бою, на якому люди виборюють своє право на самовизначення. Спільноти, що існують всередині цифрового простору і легітимність яких тривалий час заперечувалася, здобувають право висловлювати те, що репресивні суспільства відмовлялися чути. Що відбуватиметься, коли платформи, що колись постали саме для розширення простору самовираження людей, почнуть відтворювати таку ж архітектуру обмежень, яка панує в реальному світі?

Саме ця суперечність змусила кінооператора Джуда Езе поставити на своїй сторінці у Facebook питання, що цілком відображає нещодавню тенденцію: «Навіщо замінювати слово «зґвалтування» словом «примус до сексуальних дій»?». «Це потворне слово означає ще потворнішу дію», – написав Езе. «Ми повинні чітко промовляти це слово для того, щоб суспільство цілком усвідомило глибину жорстокості та нелюдяності дії, що за ним стоїть». Коментарі до його допису вибухнули розчаруванням від користувачів, чії дописи отримали спеціальні позначки, були заблоковані чи видалені, проте не тому, що вони схвально відгукнулися про насилля, а тому, що назвали речі своїми іменами та ненавмисно порушили правила платформи щодо контенту, зміст якого вважається чутливим або шкідливим. Багато користувачів – я не виняток – зазначали, що отримували суворі автоматичні попередження про те, що наші акаунти перебувають під загрозою через публікацію свідчень про сексуальне або гендерне насилля. Це явище, глобальний цифровий досвід, змусило користувачів соціальних мереж викривляти мову та висловлювання.

Не бажаючи отримати негативну реакцію, користувачі, а надто жінки, які й так перебувають у непевному становищі, тепер навмисно неправильно пишуть слово «зґвалтування», замінюють букви

зірочками, для опису використовують емодзі, або взагалі видаляють описи, де є насилля. Всі вони бояться втратити свої сторінки, які вони плакали з самого початку і які стали важливою частиною їхнього життя. Подібний словник є частиною ширшого лексикону, відомого як «алгоспик», себто набору закодованих слів, якими послуговуються користувачі задля того, щоб уникнути блокування. Такі терміни, як, наприклад, «вбити» тепер мають назву «мінуснути», «сексуальне насилля» – «СН», а «самогубство» – це «само...цид». Так само, як і «згвалтування», ці слова часом змінюють творчими написаннями або зображеннями, залежно від того, як, на думку користувачів, можна найефективніше обійти систему. Як ми до цього дійшли?

Читає машина, платить – людина

Звісно, контроль контенту виник не на порожньому місці. Його сформувала криза, що позначила перші роки існування соцмереж, коли нерегульовані цифрові простори стали родючим ґрунтом для зростання й поширення графічного насилля, кампаній гейту, вербування екстремістів та скоординованого булінгу й переслідувань. Уряди низки країн, дієвці громадянського суспільства та інші зацікавлені сторони попередили, що якщо контроль за платформами пустити на самоплин, це може посилити реальну шкоду. З огляду на це, посилився громадський тиск на контроль за соцмережами. Саме це змусило такі компанії, як Meta, що володіє Facebook та Instagram, створювати системи, здатні модерувати шкідливі слова чи контент і видаляти їх повністю, якщо виникає така потреба.

Масштаб роботи завжди був болючим питанням. Мільйони нових публікацій з'являлися щохвилини і це значно перевищувало можливості для їхньої оцінки, залишаючи модераторам близько 10 секунд для ухвалення рішення стосовно кожної публікації. Зіткнувшись із таким шаленим обсягом роботи, компанії дедалі частіше зверталися за допомогою до автоматизованих систем, розробники яких обіцяли значну швидкість і кількість обробки даних.

Те, що ми сьогодні знаємо як «алгоритми соціальних мереж» у контексті модерації контенту, є, по суті, наборами інструкцій, які дають змогу комп'ютерам виявляти закономірності та робити прогнози. На значних обсягах даних вони навчаються виявляти ознаки шкоди та працюють за допомогою класифікацій, себто процесу, що сортує контент за категоріями, такими як безпечний чи небезпечний. Такі системи потужні, працюють швидко, проте вони не здатні вловлювати нюанси чи умисел. Саме тут починаються

проблеми. Результатом їхньої роботи є машинна логіка, що звужує радикально різні види висловлювань до єдиної площини підозри. Представники компанії Мета визнали, що її автоматизовані системи часто «помиляються», бо не можуть інтерпретувати контекст так, як це роблять люди, а процедури оскарження «розчаровують своєю повільною роботою» і «не завжди дають бажаний результат». У відповідь на цю проблему технологічний гігант зараз переходить на модель Community Notes («Нотатки спільноти»), підхід, який використовує мережа X і який передає частину відповідальності користувачам, дозволяючи їм додавати коментарі до контенту.

Зменшення людського контролю підсилює вплив систем, чия здатність до інтерпретації обмежена, що збільшує ризики для спільнот, чие виживання залежить від розуміння контексту, а не його спрощення

А проте постійні помилки алгоритмічного судження не можна відокремити від більш широкої політичної економіки цифрових монополій. Великі технологічні компанії витратили роки на те, щоб скоротити робочу силу, яка могла б виправити ці помилки, замість того щоб посилювати її. Чергові хвилі звільнень персоналу вказують на те, що представники галузі прагнуть перекласти інтерпретаційні обов'язки на машини, що, в свою чергу, робить людей, здатних розуміти контекст і вловлювати нюанси, непотрібними. Хорошим прикладом є нещодавні дії TikTok. Під приводом переходу до фільтрування контенту за допомогою ШІ на початку 2025 року компанія звільнила сотні модераторів. Працівникам повідомили, що їх замінюють задля підвищення ефективності роботи мережі попри те, що автоматизовані системи часто хибно класифікують свідчення травм, політичну сатиру, висловлювання мовами меншин, а також головні новини.

Ця динаміка має значення в загальному аналізі, оскільки зменшення людського контролю підсилює вплив систем, чия здатність до інтерпретації обмежена, що збільшує ризики для спільнот, чие виживання залежить від розуміння контексту, а не його спрощення. Ніде в світі ціна цих обмежень і логіки не є такою високою, як в Африці та інших середовищах, які характеризуються значною гендерною нерівністю та соціальною нестабільністю. Тенденція, яку я називаю «позначення згвалтування іншими словами», що в інтернеті проявляється інтерпретацією страшного злочину за допомогою милих символів, є однією з найяскравіших ілюстрацій

того, як обмежувальні автоматизовані дії можуть стати інструментами шкоди. Це лінгвістичне спотворення робить насилля тривіальним актом і збільшує емоційне навантаження на тих, хто бореться за право бути почутим, а надто там, де зґвалтування систематично недооцінюють через стигматизацію, погрози та повільну роботу судової системи.

А гірше те, що мізогіни та цифрові маргінали, які добре розуміють, що евфемізми зараз прикривають шкідливу риторику, використовують цю лакуну, поширюючи прикриті свідчення про вчинення насилля як жарти чи погрози та відтворюючи насилля там, де жертви отримують попередження за те, що назвали речі своїми іменами. Якщо це реальність, то висновок має бути чесним. Не можна пом'якшувати висловлювання про зґвалтування, а тому наше завдання полягає не в тому, щоб вигадати м'якші метафори, а в тому, щоб демонтувати архітектуру, яка їх вимагає.

Важливі зміни

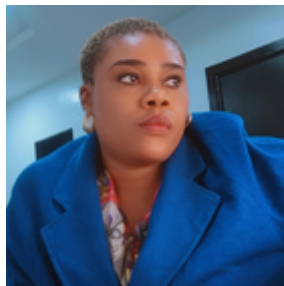
Важливі зміни починаються з матеріалістичної переорієнтації того, що платформи повинні розуміти як основне зобов'язання. Без структурної перебудови, що в центр безпеки ставить користувачів, гідність наративу та демократичну відповідальність, будь-яка реформа залишатиметься лише косметичною. Виправлення завданої шкоди також потребує повернення людського розуму в основу модерації контенту. Це передбачає наявність потужної глобальної робочої сили з навчених людей, які отримують належну оплату й психологічну підтримку, здатні ухвалювати виважені рішення без тиску квот, виконання яких не є можливим. Машини можуть допомагати, проте судити про моральність їм дозволяти не можна. Посилення захисту праці є центральним елементом цієї роботи, особливо для аутсорсингових модераторів у країнах Глобального Півдня, чиї права постійно порушуються через багаторівневі практики експлуатації субпідрядників.

Не менш важливим є перерозподіл редакційної влади. Розробка систем модерації на основі ШІ не може ґрунтуватися на даних, отриманих із промислово розвинених країн. У процесах, що визначають оцінку контенту, платформи повинні враховувати думки спільнот користувачів, жертв насилля, правозахисників та інших зацікавлених осіб. Їхній досвід потрібно враховувати при розробці наборів даних, контекстних рекомендацій та калібруванні алгоритмічних порогів. Гуманне цифрове середовище також має закріплювати право на захист свідчень. Користувачі повинні мати можливість таким чином позначати контент, аби чітко вказати, що

вони називають насилля, а не вчиняють його. Це зберігає чіткість мови, а не збільшує видалення контенту, що стирає травму.

Коли платформи переходять на модерацію, якою керує спільнота, то така зміна виглядає багатообіцяючою, адже вона повертає кінцевим користувачам можливість тлумачити контент. Проте без надійних запобіжних механізмів це може породити нові проблеми. Судження спільноти, сформовані в суспільствах, пронизаних патріархатом, гомофобією, класовими упередженнями або моральною стигматизацією, легко можуть відтворити ті самі травми, яких зазнали жертви. Будь-який перехід до нагляду, яким керує спільнота, має бути прозорим, репрезентативним і підкріпленим надійними запобіжними заходами, які не дозволять вороже налаштованій більшості визначати, чиї свідчення вважаються законними. Тільки відмовившись від евфемістичного прикриття та наполягаючи на чіткому означенні насильства, ми зможемо відновити цифровий громадянський простір як середовище, де за правду не карають, а захищають.

Переклад з англійської Ірини Савюк



Зікора Ібе

Зікора Ібе (Zikora Ibeh) — дослідниця, оглядачка, авторка подкастів і активістка, зацікавлена в соціальній і гендерній справедливості. Вона прагне змінити ситуацію в суспільстві через адвокацію публічної політики, дослідження практичних заходів та адвокацію в медіа.

