ДЕМОКРАТИЧЕСКОЕ ОБЩЕСТВО 03.12.2025 | Зикора Ибе

Нет «приукрашиванию» изнасилования

Социальные сети заставляют замолчать тех, кто открыто говорит о насилии. Там, где должна быть правда, используются эвфемизмы.

Сегодня жизнь общества в значительной степени определяют социальные сети. То, что по первоначальному замыслу было пространством развлечений и общения, превратилось в поле битвы, на котором люди борются за свое право на самоопределение. Сообщества, существующие внутри цифрового пространства и легитимность которых долгое время отрицалась, получают право выражать то, что репрессивные общества отказывались слышать. Что произойдет, когда платформы, которые когда-то возникли именно для расширения пространства самовыражения людей, начнут воспроизводить ту же архитектуру ограничений, которая царит в реальном мире?

Именно это противоречие заставило кинооператора Джуда Эзе задать на своей странице в Facebook вопрос, который полностью отражает недавнюю тенденцию: «Зачем заменять слово «изнасилование» словами «принуждение к сексуальным действиям»?». «Это уродливое слово означает еще более уродливое действие», – написал Эзе. «Мы должны четко произносить это слово для того, чтобы общество полностью осознало глубину жестокости и бесчеловечности действия, которое за ним стоит». Комментарии к его посту взорвались разочарованием от пользователей, чьи посты получили специальные пометки, были заблокированы или удалены, однако не потому, что они одобрительно отозвались о насилии, а потому, что назвали вещи своими именами и непреднамеренно нарушили правила платформы в отношении контента, содержание которого считается чувствительным или вредным. Многие пользователи – я не исключение – отмечали, что получали строгие автоматические предупреждения о том, что наши аккаунты находятся под угрозой из-за публикации свидетельств о сексуальном или гендерном насилии. Это явление, глобальный цифровой опыт, заставило

пользователей социальных сетей искажать речь и высказывания.

Не желая получить негативную реакцию, пользователи, особенно женщины, которые и так находятся в неуверенном положении, теперь намеренно неправильно пишут слово «изнасилование», заменяют буквы звездочками, для описания используют эмодзи или вообще удаляют описания, где есть насилие. Все они боятся потерять свои страницы, которые они лелеяли с самого начала и которые стали важной частью их жизни. Подобный словарь является частью более широкого лексикона, известного как «алгоспик», то есть набора закодированных слов, которыми пользуются пользователи для того, чтобы избежать блокировки. Такие термины, как, например, «убить» теперь называются «минуснуть», «сексуальное насилие» – «СН», а «самоубийство» – это «само...цид». Так же, как и «изнасилование», эти слова иногда заменяют творческими написаниями или изображениями, в зависимости от того, как, по мнению пользователей, можно наиболее эффективно обойти систему. Как мы к этому пришли?

Читает машина, платит – человек

Конечно, контроль контента возник не на пустом месте. Его сформировал кризис, который ознаменовал первые годы существования соцсетей, когда нерегулируемые цифровые пространства стали благодатной почвой для роста и распространения графического насилия, кампаний гейта, вербовки экстремистов и скоординированного буллинга и преследований. Правительства ряда стран, деятели гражданского общества и другие заинтересованные стороны предупредили, что если контроль за платформами пустить на самотек, это может усилить реальный вред. Учитывая это, усилилось общественное давление на контроль над соцсетями. Именно это заставило такие компании, как Meta, владеющая Facebook и Instagram, создавать системы, способные модерировать вредные слова или контент и удалять их полностью, если возникает такая необходимость.

Масштаб работы всегда был болезненным вопросом. Миллионы новых публикаций появлялись каждую минуту, что значительно превышало возможности для их оценки, оставляя модераторам около 10 секунд для принятия решения по каждой публикации. Столкнувшись с таким огромным объемом работы, компании все чаще обращались за помощью к автоматизированным системам, разработчики которых обещали значительную скорость и объем обработки данных.

То, что мы сегодня знаем как «алгоритмы социальных сетей» в контексте модерации контента, является, по сути, наборами инструкций, которые позволяют компьютерам выявлять закономерности и делать прогнозы. На значительных объемах данных они учатся выявлять признаки вреда и работают с помощью классификаций, то есть процесса, который сортирует контент по категориям, таким как безопасный или опасный. Такие системы мощны, работают быстро, но они не способны улавливать нюансы или умысел. Именно здесь начинаются проблемы. Результатом их работы является машинная логика, которая радикально сужает различные виды высказываний до единой плоскости подозрения. Представители компании Meta признали, что ее автоматизированные системы часто «ошибаются», потому что не могут интерпретировать контекст так, как это делают люди, а процедуры обжалования «разочаровывают своей медленной работой» и «не всегда дают желаемый результат». В ответ на эту проблему технологический гигант сейчас переходит на модель Community Notes («Заметки сообщества»), подход, который использует сеть X и который передает часть ответственности пользователям, позволяя им добавлять комментарии к контенту.

Уменьшение человеческого контроля усиливает влияние систем, чья способность к интерпретации ограничена, что увеличивает риски для сообществ, чье выживание зависит от понимания контекста, а не его упрощения

Тем не менее постоянные ошибки алгоритмического суждения нельзя отделить от более широкой политической экономики цифровых монополий. Крупные технологические компании потратили годы на то, чтобы сократить рабочую силу, которая могла бы исправить эти ошибки, вместо того чтобы усиливать ее. Очередные волны увольнений персонала указывают на то, что представители отрасли стремятся переложить интерпретационные обязанности на машины, что, в свою очередь, делает людей, способных понимать контекст и улавливать нюансы, ненужными. Хорошим примером являются недавние действия TikTok. Под предлогом перехода к фильтрации контента с помощью ИИ в начале 2025 года компания уволила сотни модераторов. Работникам сообщили, что их заменяют для повышения эффективности работы сети, несмотря на то, что автоматизированные системы часто ошибочно классифицируют свидетельства травм, политическую сатиру, высказывания

на языках меньшинств, а также главные новости.

Эта динамика имеет значение в общем анализе, поскольку уменьшение человеческого контроля усиливает влияние систем, чья способность к интерпретации ограничена, что увеличивает риски для сообществ, чье выживание зависит от понимания контекста, а не его упрощения. Нигде в мире цена этих ограничений и логики не является настолько высокой, как в Африке и других средах, которые характеризуются значительным гендерным неравенством и социальной нестабильностью. Тенденция, которую я называю «обозначение изнасилования другими словами», которая в интернете проявляется интерпретацией страшного преступления с помощью милых символов, является одной из самых ярких иллюстраций того, как ограничительные автоматизированные действия могут стать инструментами вреда. Это лингвистическое искажение делает насилие тривиальным актом и увеличивает эмоциональную нагрузку на тех, кто борется за право быть услышанным, особенно там, где изнасилование систематически недооценивают из-за стигматизации, угроз и медленной работы судебной системы.

А хуже то, что мизогины и цифровые маргиналы, которые хорошо понимают, что эвфемизмы сейчас прикрывают вредную риторику, используют эту лакуну, распространяя прикрытые свидетельства о совершении насилия как шутки или угрозы и воспроизводя насилие там, где жертвы получают предупреждения за то, что назвали вещи своими именами. Если это реальность, то вывод должен быть честным. Нельзя смягчать высказывания об изнасиловании, а потому наша задача заключается не в том, чтобы придумать более мягкие метафоры, а в том, чтобы демонтировать архитектуру, которая их требует.

Важные изменения

Важные изменения начинаются с материалистической переориентации того, что платформы должны понимать как основное обязательство. Без структурной перестройки, которая ставит в центр безопасности пользователей, достоинство нарратива и демократическую ответственность, любая реформа будет оставаться лишь косметической. Исправление нанесенного ущерба также требует возвращения человеческого разума в основу модерации контента. Это предполагает наличие мощной глобальной рабочей силы из обученных людей, которые получают надлежащую оплату и психологическую поддержку, способны принимать взвешенные

решения без давления квот, выполнение которых невозможно. Машины могут помогать, но судить о моральности им позволять нельзя. Усиление защиты труда является центральным элементом этой работы, особенно для аутсорсинговых модераторов в странах Глобального Юга, чьи права постоянно нарушаются из-за многоуровневых практик эксплуатации субподрядчиков.

Не менее важным является перераспределение редакционной власти. Разработка систем модерации на основе ИИ не может основываться на данных, полученных из промышленно развитых стран. В процессах, определяющих оценку контента, платформы должны учитывать мнения сообществ пользователей, жертв насилия, правозащитников и других заинтересованных лиц. Их опыт необходимо учитывать при разработке наборов данных, контекстных рекомендаций и калибровке алгоритмических порогов. Гуманная цифровая среда также должна закреплять право на защиту свидетельств. Пользователи должны иметь возможность таким образом помечать контент, чтобы четко указать, что они называют насилие, а не совершают его. Это сохраняет четкость языка, а не увеличивает удаление контента, что стирает травму.

Когда платформы переходят на модерацию, управляемую сообществом, то такое изменение выглядит многообещающим, ведь оно возвращает конечным пользователям возможность толковать контент. Но без надежных предохранительных механизмов это может породить новые проблемы. Суждения сообщества, сформированные в обществах, пронизанных патриархатом, гомофобией, классовыми предубеждениями или моральной стигматизацией, легко могут воспроизвести те же травмы, которые пережили жертвы. Любой переход к надзору, управляемому сообществом, должен быть прозрачным, репрезентативным и подкрепленным надежными предохранительными мерами, которые не позволят враждебно настроенному большинству определять, чьи показания считаются законными. Только отказавшись от эвфемистического прикрытия и настаивая на четком определении насилия, мы сможем восстановить цифровое гражданское пространство как среду, где за правду не наказывают, а защищают.

По этой ссылке статья откроется без VPN



Зикора Ибе

Зикора Ибе (Zikora Ibeh) – исследовательница, обозревательница, автор подкастов и активистка, заинтересованная в социальной и гендерной справедливости. Она стремится изменить ситуацию в обществе через адвокацию публичной политики, исследования практических мероприятий и адвокацию в медиа.