EDGELANDS

СОЗДАНИЕ БЕЗОПАСНЫХ ЦИФРОВЫХ ПРОСТРАНСТВ: ПРАКТИЧЕСКОЕ РУКОВОДСТВО

ВВЕДЕНИЕ

Этот документ представляет собой результат восьми недель совместной работы команды исследователь:ниц, ментор:ок и артист:ок, а также вклада приглашенных лектор:ок в рамках этапа «Pop-Down and Beyond» Института Edgelands.

Данное практическое руководство – итоговый результат проекта. Его целью является дать рекомендации по созданию безопасных цифровых пространств. Руководство основано на контент анализе первичных источников, разборе кейсов и данных опроса. Мы надеемся, что им смогут пользоваться разработчи:цы, исследовател:ьницы, художни:цы и пользовател:ьницы онлайн пространств.

Если вы хотите подробнее узнать о сложном контексте цифровой безопасности, остновных стейкхолдерах, а также о том, как определяются понятия «цифровые безопасные пространства» и «онлайн-безопасность», вы можете ознакомиться с расширенной версией этого руководства по ссылке (доступно только на английском языке).

LANDSLANDS
EDGEEDGE
LANDSLANDS

БЛАГОДАРНОСТИ

Это руководство было разработано Пулкитом Могра, Татьяной Лысовой, Лилиан Оливия Отеро, Катерин Киган, Ниной Мартин, Ммбато Оке, Джессикой МкКлирн и Джованной дэ Кустодия под руководством Нины Барановска, Даниела Одонго, Вирджинии Лаборао, Ванессы Гатеча и Лауры Гарсия Варгас.

Мы хотели бы поблагодарить всех, кто принял участие в нашем опросе и помог разработать рекомендации по созданию безопасных цифровых пространств.

Отдельная благодарность приглашённым спикерам, которые присоединились к сессиям исследовательского спринта и поделились своими знаниями и вдохновением.

Дизайн и визуальное оформление были разработаны Флавией Лоцано и Лариссой Оливэйра.

КРАТКОЕ СОДЕРЖАНИЕ

В этом руководстве собраны основные выводы, сделанные на основе контент-анализа, кейс-стади и качественного опроса. Мы предлагаем рекомендации по созданию цифровых безопасных пространств, в которых приоритет отдается эмпатии и конкретным потребностям сообществ, а не интересам платформ и извлечению прибыли.

➔ ОСНОВНАЯ ПРОБЛЕМА

В отличие от физических безопасных пространств, в цифровой среде нет четких и понятных признаков безопасности. Онлайн безопасность формируется тремя ключевыми участниками: платформами, государствами и сообществами. Однако существующие подходы остаются недостаточными, поскольку:

ПЛАТФОРМЫ создают свои сервисы, ориентируясь на условного «среднего пользователя» (как правило, цисгендерный, гетеросексуальный, белый мужчина из стран Глобального Севера со средним или высоким доходом), что делает уязвимости менее привилегированных групп невидимыми. Корпоративные интересы ставят метрики вовлеченности выше благополучия пользователей.

ГОСУДАРСТВЕННОЕ РЕГУЛИРОВАНИЕ в основном сосредоточено на предотвращении «очевидного вреда» (материалы о насилии над детьми, терроризм, мошенничество), игнорируя субъективный и контекстный характер ощущения безопасности в интернете;

СООБЩЕСТВА играют важную, но неустойчивую роль в поддержании безопасности через правила поведения и модерацию, осуществляемую волонтер:ками. Однако они остаются структурно зависимыми от архитектуры платформ.

➔ КЛЮЧЕВЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На основе анализа мы выявили четыре базовых условия, необходимых для безопасности в цифровых пространствах:

УСЛОВИЯ ВЗАИМООТНОШЕНИЙ включают уважение личных границ без необходимости конфронтации, эффективную модерацию как часть взаимоотношений (а не только обеспечение соблюдения правил), а также снижение эмоциональной нагрузки на участни:ц, которым приходится постоянно себя отстаивать и защищать.

КУЛЬТУРНЫЕ УСЛОВИЯ включают в себя социальные нормы уважения и непредвзятости, инклюзивный язык (с учетом разных языков и сообществ), а также наличие сетей взаимоподдержки, которые создают защитную среду для уязвимых групп.

ПРОЦЕДУРНЫЕ УСЛОВИЯ – это четкие и последовательные правила, возможность пользовател:ьниц самостоятельно управлять своей видимостью и сбором личных данных и легитимные модели управления, основанные на реальном опыте пользовател:ьниц, а не на произвольных корпоративных решениях.

ИНФРАСТРУКТУРНЫЕ УСЛОВИЯ касаются технической надежности и устойчивости платформ: прозрачной работы с данными, эффективных механизмов жалоб и сообщений о нарушениях, а также ответственности в случаях, когда системы дают сбой.

На основе этих условий были сформулированы основные рекомендации, разделенные на группы «обязательно должно быть», «желательно, но не обязательно» и «красные флаги», для создания безопасных цифровых пространств.

➔ ЗАКЛЮЧЕНИЕ

Это руководство служит одновременно и практическим инструментом, и приглашением переосмыслить цифровые пространства как среды, управляемые в первую очередь сообществами. В таких пространствах безопасность создается совместно самими участни:цами, а не навязывается корпоративными архитектурами, оптимизированными прежде всего под вовлеченность, а не благополучие людей. Мы также понимаем, что наше исследование имеет ограничения, такие как необходимость включения более технических перспектив, более широкого вовлечения сообществ и перевода материалов на коренные языки. В завершение мы призываем к более глубокому и продолжительному изучению вопросов онлайн безопасности.

КАК РАСПОЗНАВАТЬ, СОЗДАВАТЬ И ПОДДЕРЖИВАТЬ БЕЗОПАСНЫЕ ОНЛАЙН ПРОСТРАНСТВА: РУКОВОДСТВО

Ниже представлен практический гид по безопасным онлайн пространствам, основанный на результатах нашего исследования.

➔ ОБЯЗАТЕЛЬНЫЕ ЭЛЕМЕНТЫ

ПОЛИТИКА И ДОКУМЕНТАЦИЯ

Четкая и хорошо сформулированная документация играет ключевую роль в формировании чувства безопасности в цифровых пространствах. Уровень безопасности повышается, когда правила ясны, конкретны, доступны и видны до вступления в сообщество. Такая прозрачность позволяет людям принять осознанное решение, соответствует ли цифровое пространство их ценностям и потребностям и, следовательно, хотят ли они к нему присоединиться.

Правило должны ясно описывать допустимое поведение, нормы общения и соответствующие последствия в случае нарушений. Это снижает неопределенность и создает понимание, что вредоносное поведение не будет поощряться, что, в свою очередь, создает чувство безопасности. Важно, чтобы эти правила были написаны простым и доступным языком, понятным всем потенциальным участни:цам сообщества.

Хорошей практикой является то, когда правила сообщества прямо запрещают харассмент, травлю, язык ненависти, экстремистский контент и формы злоупотребления идентичностью (например, выдачу себя за другого человека или использование ИИ для создания неприемлемых изображений без согласия). Если формулировки правил остаются неоднозначными, должно быть пространство для их обсуждения, пересмотра и уточнения. Замена расплывчатых или абстрактный выражений, например,

таких как «экологичное» или «позитивное» общение, на конкретные ожидания (уважение, отсутствие осуждения, прозрачность) делает нормы более понятными и выполнимыми.

Правила сообщества должны разрабатываться с участием самих участни:ц сообщества. Кроме того, их необходимо регулярно пересматривать и обновлять, опираясь на реальный опыт участни:ц.

Политика управления данными также должна быть четко обозначена, уточняя как персональные данные собираются, используются, хранятся и защищаются. Этичное обращение с данными, принципы конфиденциальности по умолчанию и прозрачное, понятное объяснение мер по защите данных являются важными условиями ощущения безопасности в цифровых пространствах. Если цифровая платформа многоязычная, крайне важно, чтобы все элементы, включая информацию о защите данных, были переведены на все используемые языки.

Наконец, правила работают только тогда, когда они применяются последовательно и одинаково ко всем, включая модератор:ок и администратор:ок. Если соблюдение правил происходит выборочно или выглядит произвольным, доверие быстро разрушается, а чувство исключенности и уязвимости усиливается.

ТЕХНИЧЕСКИЕ ЭЛЕМЕНТЫ

Надежные и регулярно обновляемые инструменты безопасности являются важной частью цифровой безопасности. Они должны уважать приватность пользовател:ьниц, учитывать контекст и не быть чрезмерно карательными. К ключевыми техническим элементам, которые повышают ощущение безопасности,

относятся фильтры контента, автоматическое удаление тревожащего или незаконного контента и механизмы для жалоб и сообщений о нарушениях. Шифрование и анонимизация, защита от утечек данных и современные меры кибербезопасности, включая двухфакторную аутентификацию или систему предотвращения утечек данных, играют важную роль в формировании чувства защищенности в цифровых пространствах.

При этом автоматическая модерация иногда заходит слишком далеко, удаляя безобидный контент без учета контекста или из-за слишком жестких правил. Это может ограничить участие, подавлять легитимное самовыражение и подорвать доверие к платформе. По этой причине автоматическая модерация должна работать в связке с человеческим контролем. Также пользовател:ьницы должны иметь возможность связаться с платформой и оспорить слишком строгую автоматическую модерацию, объяснив контекст или смысл своего контента.

Техническая защита должна распространяться и на предотвращение злоупотреблений внутри сообществ, неправомерного использования личных данных и враждебной инфильтрации чувствительных или уязвимых групп. Инструменты безопасности работают эффективнее, когда они предотвращают риски заранее, а не применяются уже после того, как вред был нанесен.

Пользовател:ьницы должны иметь контроль над своей приватностью на платформе. Платформы должны предоставлять инструменты для управления видимостью профиля, выборочного раскрытия личной информации и возможности выбирать публичный или закрытый формат участия. Это дает пользовател:ьницам больше контроля над своим присутствием в интернете в зависимости от их потребностей и рисков.

Политика использования реальных имен может быть опасной в ситуациях, где анонимность и конфиденциальность критически важны для безопасности. Для многих пользовател:ьниц, особенно если они принадлежат к уязвленным или политически подавленным группам, анонимность является не предпочтением, а жизненно необходимой мерой защиты.

Наконец, платформы должны заранее и осознанно встраивать настройки и возможности для людей и сообществ с особыми потребностями. Такие возможности должны быть нормой, а не редкой привилегией или чем-то, за что пользовател:ьницам приходится постоянно бороться.

ЭЛЕМЕНТЫ СООБЩЕСТВА

При отсутствии физических границ пользовател:ьницы опираются на неформальные, но при этом очень значимые цифровые сигналы, чтобы понять, является ли пространство безопасным. Они обращают внимание на визуальные маркеры, то есть, по сути, «читают язык тела» платформы. Местоимения в описаниях профилей, инклюзивные символы, предупреждения о контенте или закрепленные вверху ленты правила поведения служат первыми признаками того, что существуют границы и что пространство осознанно поддерживается и развивается.

Эти визуальные маркеры усиливаются языковыми сигналами, которые формируют атмосферу общения внутри сообщества. Безопасные пространства чаще используют инклюзивный язык «мы», а не противопоставление «мы против них», и нередко применяют указатели тона, чтобы снизить двусмысленность. Инклюзивность, в том числе в языке, и признание разнообразия являются важными элементами правил сообщества, кодексов поведения и внутренних норм. Такие практики особенно важны для

нейродивергентных пользовател:ьниц, для которых четкая и прямолинейная коммуникация помогает снизить тревожность и избежать недопонимания.

Чтобы снизить риск нежелательной инфильтрации, сообщества могут вводить базовый процесс верификации для новых участни:ц. Например, на этапе входа можно попросить новых потенциальных участни:ц подтверждения, что они ознакомились с целями, ценностями и границами пространства, прежде чем они получат полный доступ и смогут активно участвовать в группе.

Однако важнейшим сигналом остаётся поведение. Участни:цы сообщества наблюдают за тем, как на практике работают лидерство и модерация, и оценивают безопасность по скорости, последовательности и прозрачности реакции на вред и нарушения правил сообщества. Когда язык ненависти остается без реакции или меры применяются выборочно и непоследовательно, письменные правила теряют доверие, а ощущение безопасности быстро исчезает.

В случае конфликтов внутри сообщества вместо того, чтобы сразу же прибегать к инструментам жалоб и блокировок, платформам и администратор:кам важно поощрять попытки разрешения конфликтов через уважительное и конструктивное общение, тем самым подчеркивая ответственность каждо:й участни:цы за общее пространство. При этом, если мирное разрешение невозможно, дисциплинарные меры должны применяться незамедлительно.

➔ НАШ ЧЕК-ЛИСТ РЕКОМЕНДАЦИЙ

«ОБЯЗАТЕЛЬНО ДОЛЖНО БЫТЬ» / МИНИМАЛЬНЫЕ ТРЕБОВАНИЯ

- ◯ Четкие, понятные и доступные правила и принципы сообщества, а также санкции за их нарушении
- ◯ Надежная политика защиты данных
- ◯ Ограничение возможности делать скриншоты в закрытых и приватных сообществах
- ◯ Рабочие инструменты для жалоб на вредоносный контент
- ◯ Возможность анонимного участия (по желанию пользователя)
- ◯ Контроль и подотчетность модераторо:к и администраторо:к
- ◯ Наличие модераторо:к-людей (которые имеют доступ к психологической поддержке и ресурсам, особенно при работе с травмирующим контентом)
- ◯ Использование инклюзивного языка
- ◯ Реактивные и проактивные меры против вредных высказываний и действий
- ◯ Для многоязычных платформ, корректный перевод всех элементов на все используемые языки (особенно локальные), включая пользовательские соглашения, правила поведения и пр.
- ◯ Собственные ресурсы и экспертиза в области кибербезопасности, особенно при работе с сообществами, находящимися под риском

“ЖЕЛАТЕЛЬНО, НО НЕ ОБЯЗАТЕЛЬНО”

- ◯ Проверка участни:ц на этапе вступления
- ◯ Меры по защите приватности и безопасности встроенные в архитектуру платформы
- ◯ Образовательные материалы по поведению и этикету в цифровых пространствах
- ◯ Совместное создание платформ с участием сообщества, с учетом его потребностей и предпочтений
- ◯ Модератор:ки и сотрудни:цы поддержки, которые являются частью сообщества или хорошо понимают его культурный контекст, и имеют достаточно времени и эмоциональных ресурсов для качественной помощи
- ◯ Оперативная живая поддержка, например, горячая линия, доступная в разных часовых поясах
- ◯ Специальные алгоритмические инструменты для поддержки локальных норм и языков, включая локальные языки и языки меньшинств
- ◯ Практики цифровой гигиены, например, тихие часы или ночной режим, в рамках которых модератор:ки и администратор:ки могут, например, ограничивать частоту сообщений в минуту

КРАСНЫЕ ФЛАГИ

- ◯ Непоследовательное применение правил или цензура безвредных, но отличающихся мнений
- ◯ Систематическое отсутствие реакции на вред или оправдание вреда «свободой слова»
- ◯ Обязательное раскрытие реальной личности
- ◯ «Показная» безопасность: например, формальное копирование стандартных правил
- ◯ Незащищенная техническая инфраструктура, например, уязвимость к утечкам данных и враждебной инфильтрации
- ◯ Полагание на алгоритмы или общие пользовательские соглашения для разрешения межличностных конфликтов

EDGELANDS

edgelands.institute