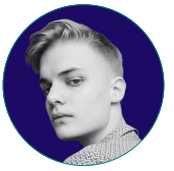
DOI: 10.14515/monitoring.2025.6.2917



И. Д. Петров

ПРЕДУПРЕЖДЕН — ЗНАЧИТ ВООРУЖЕН? АНТИВАКЦИННЫЙ КОНТЕНТ И ВОСПРИЯТИЕ ПОЛЬЗОВАТЕЛЯМИ ПОМЕТОК-ПРЕДУПРЕЖДЕНИЙ

Правильная ссылка на статью:
Петров И. Д. Предупрежден — значит вооружен? Антивакцинный контент и восприятие пользователями пометок-предупреждений // Мониторинг общественного мнения: экономические и социальные перемены. 2025. № 6. С. 110—131. https://www.doi.org/10.14515/monitoring.2025.6.2917.
For citation:
Petrov I. D. (2025) Forewarned is Forearmed? Anti-Vaccine Content and User Perception of Warning Labels. *Monitoring of Public Opinion: Economic and Social Changes.* No. 6. P. 110–131. https://www.doi.org/10.14515/monitoring.2025.6.2917. (In Russ.)

Получено: 17.02.2025. Принято к публикации: 13.10.2025.

DOI: 10.14515/monitoring.2025.6.2917

ПРЕДУПРЕЖДЕН — ЗНАЧИТ ВООРУЖЕН? АНТИВАКЦИННЫЙ КОНТЕНТ И ВОСПРИЯТИЕ ПОЛЬЗОВАТЕЛЯМИ ПОМЕТОК-ПРЕДУПРЕЖДЕНИЙ

ПЕТРОВ Игорь Дмитриевич — аспирант Департамента социологии, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, Россия
E-MAIL: idpetrov@hse.ru
https://orcid.org/0000-0002-8465-8728

Аннотация. Распространение недостоверной информации в интернете требует от социальных сетей разработки эффективных инструментов противодействия. В данной работе оценивается восприятие пользователями различных форматов интерфейсных предупреждений о недостоверном контенте на примере антивакцинных публикаций. В качестве кейса выбрана социальная сеть «ВКонтакте» — одна из самых популярных платформ в Рунете, уже имевшая опыт внедрения подобных предупреждений. Эмпирически сравниваются четыре формата предупреждений: два распространенных (блокирующее всплывающее окно и постоянный баннер) и два экспериментальных (сообщение с прямым опровержением и комбинированный вариант, разработанный с учетом пользовательских предпочтений). В рамках смешанной методологии был проведен цикл исследований, включивший полуструктурированные интервью ($N = 4$), тест предпочтений ($N = 169$) и онлайн-эксперимент ($N = 309$). Результаты показали, что пользователи статистически значимо чаще предпочитают сообщения со структурированным опровержением ложных утверждений (p-value = 0,026). В то же время ни один из форматов не привел к значимому снижению желания пользователей взаимодействовать с помеченным постом (лайк, репост, комментарий). На основе выводов исследования сформулированы прак-

FOREWARNED IS FOREARMED? ANTI-VACCINE CONTENT AND USER PERCEPTION OF WARNING LABELS

Igor D. PETROV[1] — Graduate Student at the Department of Sociology
E-MAIL: idpetrov@hse.ru
https://orcid.org/0000-0002-8465-8728

[1] HSE University, St. Petersburg, Russia

Abstract. The proliferation of misinformation online has compelled social media platforms to develop effective countermeasures. This study investigates user perceptions of different interface warning labels, using anti-vaccine content as an example. The research focuses on the social network «VKontakte» (VK), a major platform in the Russian-speaking internet segment that has previously experimented with such labels. We empirically compare four warning formats: two commonly used ones (a content-blocking interstitial pop-up and a permanent banner) and two experimental types (a refutational message and a combined format informed by user preferences). A mixed-methods approach was employed, involving a research cycle of semi-structured interviews ($N = 4$), a preference test ($N = 169$), and an online experiment ($N = 309$). The findings reveal a statistically significant user preference for messages that provide a structured refutation of false claims (p-value = 0.026). However, none of the tested warning formats resulted in a significant reduction in users' willingness to engage with the labeled post (i. e., liking, sharing, commenting). Based on these results, the study provides practical design recommendations for interface elements to counter misinformation, targeting researchers, designers, and platform developers.

DOI: 10.14515/monitoring.2025.6.2917

тические рекомендации по проектированию интерфейсных элементов для противодействия дезинформации, адресованные исследователям, дизайнерам и разработчикам платформ.

Ключевые слова: предупреждающие сообщения, вакцинация, смешанная методология, человеко-компьютерное взаимодействие, интернет-исследования, социальные сети

Keywords: warning messages, vaccination, mixed-methods research, human-computer interaction, internet research, social media

Введение

Социальные сети стали ключевым инструментом для публичного выражения мнений, коллективных дискуссий и обмена информацией. Однако их открытость и масштаб делают их также мощным каналом для распространения недостоверной информации. Одна из наиболее острых проблем в этой области — антивакцинный контент, который, маскируясь под альтернативную точку зрения, зачастую продвигает ложные утверждения о безопасности иммунизации [Benoit, Mauldin, 2021; Дудина, Сайфулина, 2023]. Учитывая, что недоверие к вакцинам признается Всемирной организацией здравоохранения одной из глобальных угроз общественному здоровью[1], поиск эффективных стратегий противодействия вакцинной дезинформации становится критически важной задачей.

В ответ на эту угрозу академическое сообщество и технологические компании активно разрабатывают стратегии, направленные на смягчение негативного воздействия дезинформации. Одной из таких стратегий стало сопровождение потенциально недостоверных публикаций предупреждающими сообщениями [Sharevski et al., 2022]. Популярные социальные сети, включая X (ранее Twitter) и Facebook*[2], начали использовать уведомления, которые блокируют часть информации в потенциально недостоверных постах, но при этом оставляют пользователям возможность ознакомиться с их содержанием[3]. Конечная цель — предоставить индивидам дополнительный контекст, который помог бы им принимать обоснованные решения насчет потребления контента [Koch, Frischlich, Lermer, 2023]. Таким образом, предупреждения в социальных сетях действуют не как запрет, а как «мягкие» подсказки (англ. soft nudges), призванные побудить пользователя к критической оценке контента, не ограничивая при этом его доступ к информации [Konstantinou, Caraban, Karapanos, 2019]. Теоретически этот механизм можно описать в рамках эвристико-систематической модели: предупрежде-

[1] Ten health issues WHO will tackle this year // World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 (дата обращения: 22.01.2025).

[2] Здесь и далее * означает: компания Meta и соцсети, которыми она владеет, признаны в России экстремистскими и запрещены.

[3] Roth Y., Pickles N. Updating our approach to misleading information // x.com. 2020. URL: https://blog.x.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information (дата обращения: 22.01.2025).

DOI: 10.14515/monitoring.2025.6.2917

ние призвано сместить обработку информации с быстрого, эвристического пути на более глубокий, систематический, способствуя принятию обоснованных решений [Koch, Frischlich, Lermer, 2023].

Эмпирические исследования подтверждают эффективность таких предупреждений в снижении доверия к фейковым новостям и намерения их распространять [Clayton et al., 2020; Mena, 2020]. Однако их успех зависит от множества факторов, включая дизайн, заметность и формат сообщения [Pennycook et al., 2020]. При этом остается открытым вопрос о том, как пользователи воспринимают различные форматы предупреждений и как это восприятие трансформируется в конкретные поведенческие интенции, такие как желание взаимодействовать с помеченным контентом. Более того, большинство исследовательских работ сфокусировано на эмпирическом изучении западных социальных сетей, тогда как специфика российского медиаполя, в частности одной из крупнейших онлайн-платформ региона «ВКонтакте», изучена недостаточно. В нашем представлении эта исследовательская прореха не только затрудняет применение человекоцентричного подхода при принятии решений об ограничении недостоверного контента, но и препятствует дальнейшему развитию стратегий противодействия дезинформации в российском контексте.

Настоящее исследование пытается восполнить данный пробел, предлагая смешанную методологию, основанную на парадигме проектно-ориентированных исследований (англ. Design Science Research, DSR). Выбор этой парадигмы обусловлен ее двойной целью: не только исследовать существующие артефакты (в нашем случае — форматы предупреждающих сообщений), но и разработать новый, улучшенный артефакт на основе выявленных пользовательских предпочтений. Такой подход обеспечивает целостную методологическую рамку, в которой этапы качественного исследования, проектирования и количественной валидации логически связаны между собой. В рамках статьи мы ищем ответы на следующие вопросы:

— Как пользователи социальной сети «ВКонтакте» воспринимают различные форматы предупреждающих сообщений?
— Какой формат предупреждающих сообщений воспринимается пользователями как наиболее предпочтительный?
— Как предупреждающие сообщения влияют на желание пользователей взаимодействовать с антивакцинными постами?

Таким образом, эмпирическая проверка строится на сравнении нескольких форматов предупреждений. Во-первых, мы оцениваем два широко распространенных формата — интерстициальный (всплывающие окна, блокирующие контент) и контекстуальный (постоянные баннеры в новостной ленте пользователя). Во-вторых, мы предлагаем и тестируем опровергающий формат сообщения, основанный на научно обоснованных принципах контраргументации (структура «Fact — Myth — Fallacy — Fact»). Наконец, чтобы проверить ценность пользовательской обратной связи, мы разрабатываем и включаем в исследование «универсальный» формат предупреждений, созданный на основе пользовательских предпочтений, выявленных на качественном этапе. Такой подход позволяет нам не только сравнить эффективность стандартных и новаторских решений, но и косвенно проверить, что сильнее влияет на поведенческие интенции: дизайн, опирающийся на теорию, или дизайн, учитывающий предпочтения пользователей.

DOI: 10.14515/monitoring.2025.6.2917

Для ответа на перечисленные вопросы мы фокусируемся на двух ключевых зависимых переменных. Первая — привлекательность предупреждения, которая позволяет оценить, насколько оно интегрируется в пользовательский опыт, не вызывая отторжения на визуальном и когнитивном уровне. Вторая переменная — намерение взаимодействовать с контентом, измеряется через готовность пользователя совершить действие, повышающее виральность поста: поставить лайк, сделать репост или прокомментировать. Именно эта поведенческая интенция напрямую связана с конечной целью предупреждений — сдерживанием распространения дезинформации через снижение вовлеченности. Анализ этих двух переменных позволяет нам не только оценивать поверхностное восприятие предупреждений, но и измерять их реальный потенциал в сдерживании распространения антивакцинного контента.

Механизм и эффективность предупреждающих сообщений

Предупреждающие сообщения представляют собой интерфейсные элементы, которые располагаются между пользователем и потенциально опасным контентом [Kaiser et al., 2021]. Основная цель таких сообщений заключается в информировании индивида, а также в создании ситуации, которая бы мотивировала его применять критическое мышление. Другими словами, использование предупреждающих сообщений можно сравнить с концепцией наджинга (англ. nudging) — техникой, направленной на подталкивание человека к определенному поведению или мыслительному процессу [Konstantinou, Caraban, Karapanos, 2019]. Эту же идею можно рассмотреть через призму эвристико-систематической модели, которая предполагает, что предупреждающее сообщение будет побуждать индивида использовать систематический путь обработки информации и, следовательно, помогать ему принимать более осознанные решения насчет потребления недостоверного контента [Koch, Frischlich, Lermer, 2023].

Академические работы подтверждают данное предположение и показывают, что предупреждающие сообщения эффективно снижают уровень доверия пользователей к сомнительной информации [Clayton et al., 2020; Porter, Wood, 2022] Например, Кэмерон Мартел и Дэвид Дж. Рэнд в своем систематическом обзоре выявили, что использование фактологического тэга «disputed» (рус. спорный / оспоренный) в постах с фейковыми новостями снижало доверие пользователей к контенту примерно на 35 % в сравнении с контрольной группой [Martel, Rand, 2023]. Кроме того, результаты смежных исследований показывают, что добавление предупреждающих сообщений может уменьшить распространение недостоверной информации путем воздействия на желание пользователей взаимодействовать с постом. Например, исследование Пола Мена показало негативную связь между наличием предупреждающего сообщения и намерением пользователей делиться недостоверным контентом [Mena, 2020]. Более того, было определено, что данная связь опосредуется воспринимаемой достоверностью, что подчеркивает комплексное воздействие предупреждений на восприятие пользователей [ibid.].

Хотя эффективность предупреждающих сообщений не вызывает сомнений, успешность их применения зависит от множества факторов. Например, важным аспектом является видимость сообщения, которая определяется его разме-

DOI: 10.14515/monitoring.2025.6.2917

ром [Gantiva et al., 2019], цветом [Silic, 2016] и местоположением на странице [Nassetta, Gross, 2020]. Исследования показывают, что незаметные предупреждения могут быть менее эффективными в снижении доверия к потенциально недостоверной информации. Этот тезис подтверждается статьей Гордона Пенникука и его коллег, которые обнаружили, что интерфейсный элемент с текстом «Disputed By 3rd Party Fact-Checker» (рус. информация оспорена в результате независимой проверки фактов), введенный Facebook* для борьбы с мифами о президентских выборах в США 2016 г., оказался малоэффективным из-за низкой заметности среди пользователей [Pennycook et al., 2020]. С помощью опроса было выявлено, что бóльшая часть пользователей даже не заметила предупреждающее сообщение рядом с опровергаемым постом [ibid.].

Помимо видимости, эффективность предупреждающих сообщений может зависеть и от специфики их содержания. Например, ряд исследований показывает, что сообщения, четко указывающие на проблему, более эффективны, чем общие и недетализированные надписи [Clayton et al., 2020; Epstein et al., 2022]. В частности, метка «false» (рус. ложный) может быть менее эффективной, чем пояснение, разъясняющее конкретное заблуждение [Clayton et al., 2020]. Здесь важно отметить, что иногда предупреждения, делающие категоричные выводы о содержании потенциально опасного поста, могут быть неуместны, так как даже нежелательная информация может оказаться правдивой [Allen, Watts, Rand, 2024]. Например, посты противников вакцинации часто упоминают возможные побочные эффекты иммунизации. Поскольку эта информация технически не является ложной, отмечать такие посты меткой «false» (рус. ложный) будет неправильным [ibid.]. Эффективнее будет подчеркнуть, что преимущества вакцинации значительно перевешивают риск развития осложнений.

Несмотря на многообразие способов представления предупреждающих сообщений, современный ландшафт социальных медиа характеризуется доминированием интерстициальных и контекстуальных сообщений [Kaiser et al., 2021]. Интерстициальные сообщения представляют собой баннеры или всплывающие окна, которые блокируют часть информации от просмотра и требуют от пользователя выполнить определенное действие [Sharevski et al., 2022]. Например, подтвердить свое ознакомление с тем, что контент может содержать недостоверную информацию [ibid.]. Контекстуальные сообщения, в свою очередь, встраиваются в контент и напоминают пользователю о правилах или ограничениях в определенной публикации [Guo et al., 2024]. Это могут быть предупреждения о содержании отталкивающих сцен или о необходимости проверить достоверность информации перед ее распространением [ibid.]. Исследования показывают, что все названные способы предупреждения воспринимаются пользователями позитивно [Akhawe, Felt, 2013; Xie et al., 2022]. Например, Цзиньи Се и ее коллеги изучили три формата предупреждающих сообщений (интерстициальные, контекстуальные и подсвечивающие[4]) и пришли к выводу, что интерстициальные предупреждающие сообщения воспринимаются пользователями как наиболее уместные с точки зрения баланса между эффективным

[4] В рамках изучаемой статьи подсвечивающее предупреждение понимается как выделение в интерфейсе определенного фрагмента текста с недостоверной информацией и размещение рядом с ним пояснения о возможной ошибочности данных.

DOI: 10.14515/monitoring.2025.6.2917

предупреждением и относительной ненавязчивостью по сравнению с подсвечивающим форматом, который часто считался слишком агрессивным [Xie et al., 2022].

В российском медиапространстве предупреждающие сообщения не получили такого же широкого распространения, как на западных платформах. Исключение составляет социальная сеть «ВКонтакте», которая в 2019 г. начала сопровождать антивакцинные сообщества предупреждением о возможной недостоверности информации [Petrov, 2022]. Однако данное вмешательство не нашло своего распространения и к 2025 г. всплывающее окно с предупреждением перестало появляться даже в крупных антивакцинных сообществах, аудитория которых превышает 50 тыс. участников[5]. При этом официальных заявлений от администрации «ВКонтакте» с объяснением причин прекращения этой практики не публиковалось. В связи с этим можно сделать вывод, что в локальном информационном пространстве по-прежнему существует значительный пробел в реализации мер по предупреждению пользователей о потенциально опасной информации.

Принципы опровержения

В процессе опровержения не всегда достаточно предоставить правдивую информацию, также важно убедиться, что она сохранится в сознании индивида [Johnson, Seifert, 1994]. Способ представления опровергающего нарратива играет особенно важную роль, так как от него зависит, какую информацию человек запомнит и примет. Например, излишнее внимание к заблуждению может привести к тому, что в памяти отложится именно ложная информация [Lewandowsky et al., 2012]. В литературе это явление называется обратным эффектом (англ. backfire effect) и объясняется тем, что каждый раз, когда люди слышат или читают опровергаемое утверждение, они становятся с ним более знакомыми [Swire-Thompson, DeGutis, Lazer, 2020]. Это уменьшает когнитивные усилия, необходимые для обработки ложного утверждения, а следовательно, повышает вероятность того, что люди поверят в его правдивость (эффект иллюзии правды) [Hassan, Barber, 2021]. Данную ситуацию можно проиллюстрировать исследованием Сары Плувиано и ее коллег, которое показало, что повторение мифов о вакцинах и их последующее опровержение научными фактами вызывало парадоксальный эффект: ложные убеждения усиливались по сравнению с контрольной группой [Pluviano, Watt, Della Sala, 2017]. Данные результаты не говорят о том, что опровержение неминуемо приводит к закреплению недостоверной информации, а скорее дают понять, что оспаривание — это сложный процесс, к которому стоит подходить с умом.

Одним из наиболее известных способов борьбы с обратным эффектом является обрамление недостоверной информации фактами. Другими словами, для увеличения вероятности принятия правдивой информации необходимо вставлять ложную информацию между несколькими истинными высказываниями. Джордж Лакофф называет такую технику «сэндвич правды» (англ. truth sandwich) и предлагает использовать ее для более эффективного написания новостных статей, которые рассказывают о тех или иных логических заблуждениях[6]. Ряд эксперименталь-

[5] Примеры сообществ автор готов предоставить по запросу.

[6] The Truth Sandwich: A Better Way to Mythbust // Communicate Health. 2020. URL: https://communicatehealth.com/wehearthealthliteracy/the-truth-sandwich-a-better-way-to-mythbust (дата обращения: 19.12.2024).

DOI: 10.14515/monitoring.2025.6.2917

ных исследований указывает на то, что данный подход действительно эффективен и может применяться не только в журналистике [Kotz, Giese, König, 2023; Tulin et al., 2024]. Например, в исследовании Лауры Кёниг было показано, что представление мифа о питании в названном формате приводило к тому, что миф воспринимался участниками как менее достоверный [König, 2023].

Расширенная версия «сэндвича правды» — структура «Fact — Myth — Fallacy — Fact», предложенная Джоном Куком, Стефаном Левандовски и их коллегами [7]. В частности, в своей книге «The Debunking Handbook» авторы утверждают, что эффективное опровержение мифов должно состоять из четырех шагов: (1) четкое указание факта, (2) идентификация мифа, (3) указание ошибки и (4) повторное утверждение факта. Стоит отметить, что данный способ опровержения основан на принципе научного консенсуса, что делает его особенно весомым в практических задачах по борьбе с дезинформацией.

В целом можно констатировать, что предупреждающие сообщения являются перспективным инструментом противодействия недостоверной информации в социальных сетях. Для достижения максимального эффекта предупреждения должны быть хорошо заметны, специфичны (т. е. четко указывать на недостоверный элемент и содержать факты) и доступны для понимания. На основании этой информации мы выдвигаем следующие гипотезы:

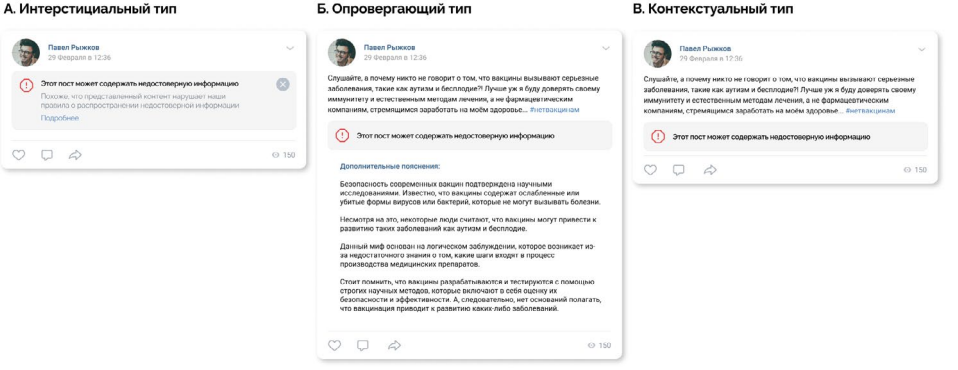
Гипотеза 1: включение предупреждающего сообщения к посту с недостоверной информацией уменьшит вероятность взаимодействия с ним.

Гипотеза 2: предупреждающие сообщения с аргументативными пояснениями будут более эффективны в уменьшении вероятности взаимодействия с постом, чем другие форматы предупреждающих сообщений.

Первичный прототип предупреждающего сообщения

На основании выводов, полученных из литературного обзора, были сформированы три формата предупреждающих сообщений (см. рис. 1).

Рис. 1. Первичные типы предупреждающих сообщений



[7] Lewandowsky S., Cook J., Ecker U. K. H., Albarracín D., Amazeen M. A., Kendeou P., Lombardi D., Newman E. J., Pennycook G., Porter E. Rand D. G., Rapp D. N., Reifler J., Roozenbeek J., Schmid P., Seifert C. M., Sinatra G. M., Swire-Thompson B., van der Linden S., Vraga E. K., Wood T. J., Zaragoza M. S. The Debunking Handbook // Bigital Commons. 2020. URL: https://digitalcommons.unl.edu/scholcom/245 (дата обращения: 12.01.2025).

DOI: 10.14515/monitoring.2025.6.2917

Два из трех форматов предупреждающих сообщений (интерстициальный и контекстуальный) представляют собой широко известные способы предупреждения, которые интегрированы в такие социальные сети, как X, Facebook*, Instagram* и YouTube [Guo et al., 2024]. Опровергающий формат базируется на ранее упомянутых принципах опровержения и предлагается в качестве тематической (англ. topic-aware) альтернативы, которая адресует конкретные заблуждения противников вакцинации. Все из названных предупреждений для дизайна/эксперимента исследования были спроектированы в соответствии с дизайн-решениями, которые применяются разработчиками социальной сети «ВКонтакте». В частности, во время формирования прототипов мы опирались на открытую библиотеку компонентов VK UI, что помогло не только создать предупреждающие сообщения, но и включить их в реплицированную (воссозданную) версию новостной ленты «ВКонтакте».

Интерстициальный формат. Данное предупреждение блокирует потенциально опасную информацию от просмотра, однако оставляет пользователям возможность ознакомиться с постом путем нажатия на соответствующую кнопку. Как видно из рисунка 1(А), помимо уведомления, закрывающего контент, в данном предупреждающем сообщении также присутствует кнопка «Подробнее», которая позволяет пользователю получить подробное описание причин появления данного уведомления.

Опровергающий формат. Данное предупреждение представляет собой модифицированную версию контекстуального формата, включающую постоянное (неизменное) окно с тематическим пояснением, направленным на опровержение распространенных заблуждений о вакцинации. Опровергающее сообщение, как показано на рисунке 1(Б), основано на четырехступенчатом принципе контраргументации, разработанном Джоном Куком, Стефаном Левандовски и их коллегами. Применение этого принципа может быть проиллюстрировано следующим образом.

1. Факт (Fact). Безопасность современных вакцин подтверждена научными исследованиями. Известно, что вакцины содержат ослабленные или убитые формы вирусов или бактерий, которые не могут вызывать болезни.

2. Миф (Myth). Несмотря на это, некоторые люди считают, что вакцины могут привести к развитию таких заболеваний, как аутизм и бесплодие.

3. Заблуждение (Fallacy). Данный миф основан на логическом заблуждении, которое возникает из-за недостаточного знания о том, какие шаги входят в процесс производства медицинских препаратов.

4. Факт (Fact). Стоит помнить, что вакцины разрабатываются и тестируются с помощью строгих научных методов, которые включают в себя оценку их безопасности и эффективности. Следовательно, нет оснований полагать, что вакцинация приводит к развитию каких-либо заболеваний.

Контекстуальный формат. С точки зрения функциональности это минимально интерактивный элемент интерфейса, он не блокируют контент и не предоставляет пользователю дополнительной информации. Он отличается от других форматов предупреждений тем, что не имеет опции закрытия и остается постоянным элементом на экране. Такой подход обеспечивает постоянное напоминание пользователю о том, что пост может содержать потенциально опасную информацию. Контекстуальное предупреждение представлено на рисунке 1(В).

DOI: 10.14515/monitoring.2025.6.2917

Методология

Работа выполнена в рамках парадигмы проектно-ориентированных исследований (англ. Design Science Research, DSR), которая ориентирована на создание и эмпирическую оценку артефактов, направленных на решение конкретных практических задач. В нашем случае таким артефактом стал дизайн предупреждающих сообщений о недостоверной информации. Для достижения поставленных целей был применен смешанный методологический подход с последовательной объяснительной стратегией (англ. exploratory sequential design). Этот подход предусматривает первичный сбор и анализ качественных данных для выработки обоснованных требований к проектированию артефакта, за которым следует количественная проверка на расширенной выборке.

Дизайн исследования

Исследование состояло из трех последовательных фаз, каждая из которых решала конкретные задачи в рамках DSR-парадигмы.

Качественная фаза: проведение полуструктурированных интервью с целью выявления паттернов восприятия, языковых конструкций и потенциальных барьеров, связанных с различными форматами предупреждений. Результаты этой фазы послужили основой для проектирования нового артефакта — «универсального» формата предупреждающего сообщения, интегрирующего выявленные пользовательские предпочтения.

Количественная фаза 1: количественный тест предпочтений для статистической оценки того, какой из четырех форматов сообщений (три исходных + один спроектированный) пользователи считают предпочтительным. Эта фаза позволила верифицировать инсайты, полученные на качественном этапе, на более широкой выборке.

Количественная фаза 2: межгрупповой онлайн-эксперимент для тестирования гипотез о влиянии предупреждающих сообщений на ключевую поведенческую интенцию пользователя — желание взаимодействовать с помеченным контентом (лайк, репост, комментарий).

Участники и процедура сбора данных

Этап 1: полуструктурированные интервью. На разведывательном этапе нашего исследования были проведены четыре полуструктурированных интервью. Их общая продолжительность составила 2 часа и 20 минут (в среднем 35 минут на каждое интервью). Все опрашиваемые были найдены с использованием невероятностной выборочной стратегии, а именно методом «снежного кома» с одной точкой входа. Данный подход распространен в области исследований пользовательского опыта (UX) и считается уместным на первичных этапах работы, проводимой в рамках парадигмы DSR. Выборка включала одного мужчину и трех женщин в возрасте от 19 до 40 лет с различным образовательным и профессиональным бэкграундом.

Основная же часть интервью проводилась в соответствии с интервью-гайдом, состоящим из 45 вопросов[8] и разделенным на три блока: (1) социально-демогра-

[8] В данном исследовании интервью носят не этнографический, а целенаправленный (англ. focused) характер, что типично для этапа проектирования в UX-исследованиях. Их цель — получить обратную связь по конкретным интерфейсным решениям, а не всесторонние биографические нарративы.

DOI: 10.14515/monitoring.2025.6.2917

фический блок, (2) вопросы о предыдущем опыте взаимодействия с антивакцинной информацией, а также об отношении пользователей к контролю за распространением потенциально недостоверного контента; (3) вопросы о предупреждающих сообщениях. В третьем блоке информантам были показаны три интерактивных прототипа с предупреждающими сообщениями, изображенными на рисунке 1. В данной части интервью также применялась техника «думай вслух» (англ. think aloud protocol), в которой информантов просили описывать свои действия, мысли и эмоции, возникающие во время взаимодействия с новостной лентой, в частности с предупреждающим сообщением. Использование данной техники обусловлено ее признанием в кругах специалистов по UX, а также возможностью получить ценные инсайты о том, как улучшить эффективность элементов интерфейса и сделать их более понятными для пользователей.

Этап 2: тест предпочтений. В исследовании приняли участие 259 человек (средний возраст 35 лет, SD = 7,90), отобранных через краудсорсинговую платформу Pathway. Среди них мужчины составили 52 %, женщины — 47 %, предпочли не указывать пол — 1 %. На этапе анализа 90 участников были отфильтрованы за неправильные ответы на контрольный вопрос[9]. В результате анализировалась информация по 169 респондентам (средний возраст 36 лет, SD = 7,57): 52 % мужчины, 47 % женщины, 1 % предпочли не указывать пол.

В рамках тестирования участникам было предложено выбрать из четырех форматов предупреждающих сообщений тот, который им нравится больше всего. Предупреждающие сообщения были представлены в виде статичных картинок, демонстрирующих все функциональные раскрытия сообщения (предупреждение после нажатия на кнопку «подробнее», предупреждение после нажатия на кнопку закрытия) — чтобы убедиться в том, что респонденты имеют целостное представление о содержании плашек. После выбора сообщения участников просили заполнить анкету, состоящую из семи вопросов (три открытых и четыре закрытых). Ознакомиться с собранными данными можно в публичном репозитории GitHub[10].

Этап 3: онлайн-эксперимент. В эксперименте приняли участие 350 человек (средний возраст 37 лет, SD = 7,76), отобранных через краудсорсинговую платформу Pathway. Среди них: мужчины — 50 %, женщины — 49 %, предпочли не указывать пол — 1 %. На этапе анализа 41 участник был отфильтрован за неправильные ответы на контрольный вопрос. В результате анализировалась информация по 309 респондентам (средний возраст 37 лет, SD = 7,85): 50 % — мужчины, 49 % — женщины, 1 % не указали пол.

Перед началом тестирования все участники были случайным образом разделены на пять экспериментальных групп. Четырем группам был показан интерактивный прототип новостной ленты «ВКонтакте» с одним из вышеописанных предупреждающих сообщений, в то время как пятой (контрольной) группе был продемонстрирован прототип без предупреждающего сообщения. До фильтрации по ответу на контрольный вопрос[11] в каждой группе было по 70 участников, после фильтрации это значение изменилось следующим образом: интерстициаль-

[9] Контрольный вопрос № 1: «Сколько предупреждающих сообщений Вам было представлено?»

[10] Ссылка на репозиторий. URL: https://github.com/nirs-paper/warnings-nirs-paper.

[11] Контрольный вопрос № 2: «Сколько постов в новостной ленте вам было показано?»

DOI: 10.14515/monitoring.2025.6.2917

ная группа ($n = 62$), контекстуальная группа ($n = 62$), опровергающая группа ($n = 64$), универсальная группа ($n = 63$), контрольная группа ($n = 58$). В рамках эксперимента участников просили внимательно изучить представленный прототип, после чего задавали ряд закрытых вопросов об антивакцинном посте. В частности, основные вопросы касались желания участников поставить лайк, оставить комментарий и сделать репост. Каждая из этих переменных была представлена в виде шкалы Ликерта, где 1 означало «совершенно не согласен», а 7 — «полностью согласен».

Анализ данных

Этап 1. Аудиозаписи полуструктурированных интервью были расшифрованы с помощью Teamlogs, проверены и проанализированы методом индуктивного тематического кодирования для выявления повторяющихся тем.

Этап 2. Для проверки гипотезы о том, что наблюдаемая вероятность выбора того или иного формата предупреждающих сообщений отличается от 0,25, был проведен двухсторонний биномиальный тест. В данном случае 0,25 используется из предположения, что при отсутствии предпочтения выбор любого из четырех форматов предупреждающих сообщений будет равновероятен. Помимо вышеупомянутого анализа, также были проведены три теста на независимость (Критерий χ^2 Пирсона), проверяющие наличие статистически значимых ассоциаций между предпочитаемым форматом предупреждающего сообщения и следующими переменными: (1) пол участника, (2) наличие предыдущего опыта взаимодействия с предупреждающими сообщениями и (3) согласие с утверждением о том, что посты с потенциально недостоверной информацией должны помечаться предупреждениями. Для упрощения задачи все вычисления были проведены с помощью пакета stats в RStudio.

Для анализа причин, по которым пользователи предпочитали тот или иной вариант, использовалась встроенная в платформу Pathway функциональность для проведения тематического анализа качественных данных. Полученные выводы были проверены, а затем подкреплены открытыми ответами участников.

Этап 3. Учитывая, что шкала Ликерта может быть рассмотрена как ординальная аппроксимация непрерывной переменной, статистический анализ, использующий эту шкалу, может быть выполнен различными способами [Sullivan, Artino, 2013]. В данном случае проверка наличия статистически значимых различий в желании пользователей взаимодействовать с антивакцинной публикацией между группами, получавшими разные типы предупреждающих сообщений, была проведена двумя способами: через дисперсионный анализ и тест X^2 Пирсона. В первом случае были использованы параметрическая и непараметрическая версии ANOVA, а также проверена выполнимость статистических предположений о нормальности распределения (тест Шапиро — Уилка) и гомогенности дисперсии (тест Левена). Во втором случае использовалась классическая версия теста, однако целевые переменные были соединены в укрупненные категории.

Результаты исследования

Анализ интервью показывает, что предупреждающие сообщения действительно повышают визуальную заметность потенциально недостоверного контента, особенно в случае опровергающего типа:

DOI: 10.14515/monitoring.2025.6.2917

Ну, я сначала подумала о том, что из поста в четыре строчки сделали огромный пост, и, естественно, он привлечет больше внимания, чем изначально. (Инф. 3, Женщина, 27 лет, высшее образование)

Однако исследование выявило, что повышенная заметность не всегда приводит к взаимодействию с контентом. Большинство пользователей заявили, что склонны игнорировать подобные публикации из-за заранее сформированной позиции, которая не зависит от содержания или наличия предупреждения:

Потому что, если я считаю, что вакцинация для меня — это польза, то я не открываю эти посты и не читаю их даже. Я, конечно, могу что-то про них слышать, но так, чтобы читать и интересоваться, — нет. (Инф. 2, Женщина, 40 лет, среднее профессиональное образование)

Желание ознакомиться с помеченным постом также может зависеть от его тематики, на что обратили внимание два информанта:

Если бы я увидела, что там что-то про вакцинацию, я бы, наверное, просто пролистнула, потому что мне это не особо интересно. Если бы там было написано, что контент просто содержит недостоверную информацию, я чисто из любопытства открыла бы и посмотрела, что же там такое интересное. (Инф. 3, Женщина, 27 лет, высшее образование)

Другим интересным наблюдением является то, что ни один из информантов не выбрал контекстуальный формат предупреждения в качестве предпочтительного. Трое из четырех отдали предпочтение интерстициальному формату, в то время как четвертый предпочел опровергающий тип, отметив, что дополнительные пояснения должны быть скрыты в раскрывающееся поле.

Мне кажется, первый [интерстициальный тип], *потому что, как я понял, что если нажать на крестик, то там окажется сам пост. Плюс там есть кнопка «Подробнее», которую можно нажать и ознакомиться с пояснением к самому сообщению.* (Инф. 1, Мужчина, 19 лет, среднее профессиональное образование)

Участники также выразили обеспокоенность этическими последствиями чрезмерного сокрытия контента, указав на риски ограничения свободы выражения и формирования одностороннего информационного поля:

Ну, я считаю, что в принципе какое-то ограничение на распространение информации — это не очень хорошо. В любом случае это [предупреждающее сообщение] *регулируется только какой-то определенной стороной, которая имеет свое субъективное мнение. Получается, что тогда пользователи социальных сетей будут иметь только возможность ознакомиться с одной точкой зрения, а это неправильно.* (Инф. 4, Женщина, 24 года, высшее образование)

DOI: 10.14515/monitoring.2025.6.2917

Кроме того, информанты последовательно отмечали недостаточную прозрачность критериев, по которым социальные сети маркируют сообщения как недостоверные. Пользователи хотят видеть ссылки на исследования, регламенты или другие верифицируемые основания, что могло бы повысить доверие к механизмам модерации:

> *В идеальном мире здесь, конечно, бы какую-нибудь дать ссылочку, не знаю, на регламент о том по каким критериям они помечают недостоверный контент. <…> Ну, короче, суть в том, что я бы, наверное, дополнительное объяснение свернула бы в «подробнее» и дала бы пояснения о том, почему такие выводы, и так далее.* (Инф. 3, Женщина, 27 лет, высшее образование)[12]

Наконец, информанты указали на вероятность ошибок алгоритмов, способных помечать корректные публикации как сомнительные, и подчеркнули необходимость внедрения инструмента для обжалования таких решений, чтобы повысить справедливость и точность системы:

> *Я бы здесь еще добавила: «Если вы не согласны с тем, что сообщение содержит недостоверную информацию, то пожалуйтесь на него». Не знаю, ну что-то такое. Потому что сталкивалась с тем, что какой-то пост, не знаю, про банановые панкейки маркировали предупреждением. То есть ничего такого там не было, и мне от этого даже как-то обидно было — человек постарался, написал* [пост]*, а его текст по ошибке прикрыли.* (Инф. 4, Женщина, 24 года, высшее образование)

Формирование универсального типа предупреждающих сообщений

На основании выводов, полученных из интервью, были сформулированы следующие требования к четвертому (универсальному) типу предупреждающих сообщений, который непосредственно учитывает основные замечания и пожелания пользователей:

— Оставить интерстициальную блокировку контента.

— Добавить в предупреждающее сообщение ссылку на регламент, согласно которому социальная сеть отмечает посты плашками.

— Добавить в заголовок предупреждающего сообщения тематику поста.

— Добавить возможность скрытия дополнительные пояснения в отдельное окошко. В нем также стоит сделать отсылку на научный источник информации.

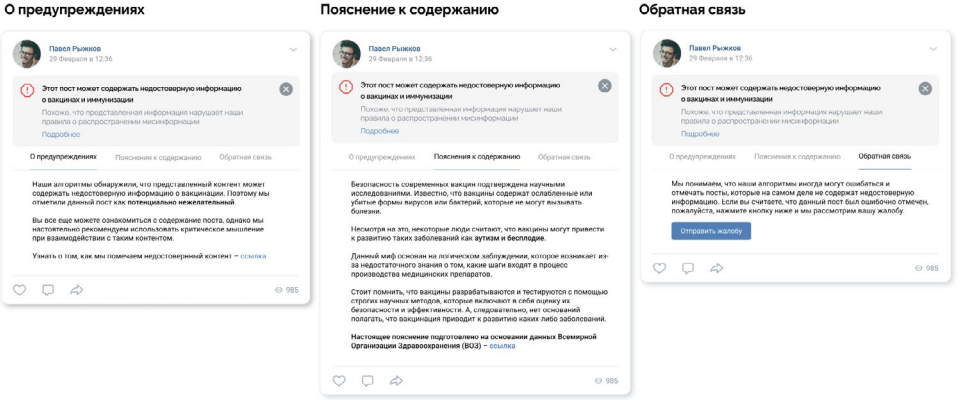
— Добавить возможность отправки жалобы на предупреждающее сообщение на случай, если оно ошибочно.

Предупреждение, разработанное на основе этих принципов, показано на рисунке 2. Оно представляет собой комбинацию интерстициального и опровергающего типа, дополненную рядом улучшений. Например, к кнопке «Подробнее» была добавлена композитная структура, что позволило включить дополнительную информацию, сохраняя при этом относительную компактность сообщения.

[12] Здесь и далее в цитатах знак <…> обозначает фрагмент, намеренно опущенный автором исследования. Он используется для сокращения цитаты с сохранением её основного смысла, удаления повторов или нерелевантных для конкретного контекста деталей.

DOI: 10.14515/monitoring.2025.6.2917

Рис. 2. Универсальный тип предупреждающего сообщения



Тест предпочтений

Результаты точного биномиального теста показывают, что наблюдаемая вероятность выбора пользователями опровергающего типа в качестве предпочтительного статистически значимо отличается от 0,25 (при доверительном уровне в 0,05), то есть именно этот способ представления предупреждающего сообщения наиболее привлекателен в глазах пользователей. Общее описание результатов теста для каждого типа предупреждающих сообщений представлено в таблице 1.

Таблица 1. ***Результаты биноминального теста предпочтений***

Формат сообщения	Количество респондентов, выбравших данный тип	p-value	Наблюдаемая вероятность выбора	95% CI
Интерстициальный	37 (22%)	0,3755	0,21	[0,160, 0,289]
Контекстуальный	37 (22%)	0,3755	0,21	[0,160, 0,289]
Опровергающий	55 (33%)	0,0263	0,33	[0,255, 0,401]
Универсальный	40 (23%)	0,7233	0,24	[0,175, 0,308]

Учитывая то, что статистическая значимость не дает информации о силе взаимосвязи, в эмпирических исследованиях также рекомендуется рассчитывать размер эффекта [Fritz, Morris, Richler, 2012]. В данном случае была вычислена метрика Cohen's h, которая позволяет оценить величину различия между двумя вероятностями [ibid.]. Она рассчитывается по следующей формуле, где *p* и *p0* — это наблюдаемая и ожидаемая вероятность, соответственно:

$$h = 2(\arcsin(\sqrt{p}) - \arcsin(\sqrt{p_0})).$$

После подстановки значений, полученных из биноминального теста (опровергающий тип), можно сказать, что размер эффекта составляет ~0,18. Это говорит о том, что, хотя статистический анализ показывает значимые результаты, предпочтение к опровергающему типу незначительно.

DOI: 10.14515/monitoring.2025.6.2917

Результаты проведенных тестов на независимость показывают, что предпочтения к тому или иному типу предупреждающих сообщений не связаны с полом (p-value = 0,430), предыдущим опытом взаимодействия с предупреждениями (p-value = 0,135) и согласием с тем, что пометка потенциально недостоверных постов необходима (p-value = 0,107). Другими словами, выявить статистически значимых различий в предпочтениях между разными группами пользователей не удалось.

О причинах выбора

Качественный анализ открытых ответов из теста предпочтений выявил четкую систему аргументации, которой респонденты руководствовались при выборе предпочитаемого типа предупреждающих сообщений. Аргументы были тесно связаны с воспринимаемой эффективностью, лаконичностью и уважением к агентности пользователя.

Интерстициальный тип был выбран респондентами преимущественно благодаря наличию краткого предупреждения, предваряющего основной контент и разъясняющего причины его маркировки. Ключевыми факторами эффективности признаны краткость и ясность сообщения, в противовес большим текстам, которые, по мнению пользователей, чаще игнорируются. Как отметил один из респондентов, *«Краткое и понятное предупреждение лучше подходит! Длинное никто не будет читать до конца, слишком долго, нудно, подумают — ерунда какая-то. К тому же, если предупреждение длиннее самого поста, то это как-то вообще не очень смотрится»* (респ. 127, Мужчина, 40 лет).

Контекстуальный тип привлек пользователей своей лаконичностью, понятностью и минималистичным дизайном, который не перегружал визуальное пространство. Респонденты подчеркивали, что данная интервенция, будучи лаконичной и визуально ненавязчивой, эффективно выполняла функцию предупреждения, не затрудняя доступ к первоисточнику и не навязывая конкретную оценку. Этот подход рассматривался как компромиссный, обеспечивающий информирование пользователя без избыточного контроля над контентом: *«Полностью блокировать информацию нельзя, так как каждый сам в праве решать, какую информацию потреблять и чему верить, поэтому один из вариантов отметаем сразу. В остальных вариантах слишком много пояснительной информации, поэтому выбираем самый простой, без лишней воды»* (респ. 87, Мужчина, 33 года).

Опровергающий тип был выбран теми респондентами, которые видели ценность в предоставлении развернутой аргументации и дополнительного контекста. Данный подход воспринимался как наиболее уважающий право пользователя на самостоятельный анализ и формирование выводов на основе конкурирующих точек зрения. Эта позиция иллюстрируется следующим высказыванием: *«Подобный пост не следует скрывать. Он не несет в себе экстремизма. Это точка зрения. Должна быть возможность ознакомиться с ней. Иначе это еще больше разожжет у читателя интерес к этому посту. А уже под постом должна быть альтернативная точка зрения. А читатель уже сам решает, чему верить»* (респ. 187, Мужчина, 37 лет).

Универсальный тип был охарактеризован респондентами как наиболее комплексный и информативный. Назывались такие его преимущества, как детальное объяснение причин блокировки контента и наличие механизма обратной связи.

DOI: 10.14515/monitoring.2025.6.2917

В целом респонденты оценили этот вариант как способствующий более глубокому осмыслению информации: *«Идеальный вариант, так как поясняется почему выскочило предупреждение, есть краткий экскурс о том, о чем говорится в посте и обратная связь. Заставит читателя помозговать, осмыслить, и возможно, остудит троллей интернетных»* (респ. 143, Женщина, 35 лет).

Результаты онлайн-эксперимента

Результаты однофакторного дисперсионного анализа (см. табл. 2) указывают на отсутствие статистически значимых различий в среднем желании взаимодействовать с постом через лайк, комментарий или репост между экспериментальными группами.

Таблица 2. ***Результаты дисперсионного анализа***

Переменная	F-статистика	p-value	Значение χ^2	95 % CI
Желание поставить лайк	0,57	0,687	7,40e-03	[0,00, 1,00]
Желание сделать репост	0,23	0,921	7,40e-03	[0,00, 1,00]
Желание оставить комментарий	0,26	0,906	3,35e-03	[0,00, 1,00]

Результаты непараметрической версии ANOVA (критерий Краскела — Уоллиса) подтверждают данные выводы, указывая на отсутствие статистически значимых различий между группами (p-value > 0,05). Как и в случае с ANOVA, тест хи-квадрат не позволил выявить статистически значимой связи между желанием взаимодействовать с постом и экспериментальными группами (см. табл. 3).

Таблица 3. ***Результаты теста χ^2***

Переменная	χ^2	p-value
Желание поставить лайк	4,2574	0,8332
Желание сделать репост	9,4266	0,3076
Желание оставить комментарий	2,7284	0,9502

Заключение

В работе проведен анализ пользовательского восприятия предупреждающих сообщений. Это позволило выявить как слабые, так и сильные стороны каждого из четырех способов визуального представления предостерегающих уведомлений (интерстициальные, контекстуальные, опровергающие и универсальные), а также подтвердить наличие статистически значимого предпочтения в отношении опровергающего формата сообщений (p-value = 0,02631, Cohen's h = 0,1767). Это дает основания полагать, что русскоязычные пользователи позитивно воспринимают дополнительные пояснения, представленные в форме так называемого «сэндвича правды», где недостоверная информация обрамляется истинными высказываниями. Более того, исследование показало, что включение предупреждающего

DOI: 10.14515/monitoring.2025.6.2917

сообщения не снижает вероятность взаимодействия с постом через лайк, репост или комментарий, и это противоречит выводам, сделанным в предыдущих исследованиях (например, [Clayton et al., 2020]). Несмотря на это, мы не считаем, что данный результат свидетельствует о неэффективности предупреждающих сообщений. Отсутствие статистически значимой ассоциации, вероятно, объясняется двумя ключевыми факторами. Во-первых, высокий уровень образования респондентов в выборке: 65 % участников обладают степенью бакалавра или выше, что предполагает развитые навыки критического мышления и, возможно, изначально более скептическое отношение к антивакцинному контенту. Для такой аудитории базовый уровень доверия к ненадежным источникам мог быть изначально низким, что ограничило пространство для наблюдаемого эффекта от предупреждения. Во-вторых, специфика методологии: в ходе эксперимента мы напрямую спрашивали респондентов об их желании взаимодействовать с постом. Эта прямая постановка вопроса могла спровоцировать социально одобряемые ответы, сместив результаты в сторону большей осторожности, независимо от реальных намерений участников. Вероятно, лучшей стратегией было бы отследить взаимодействие с антивакцинным постом через целевые действия пользователей.

Также для более комплексной оценки эффективности предупреждений в будущих исследованиях необходимо целенаправленно изучить их влияние на два ключевых поведенческих аспекта: на желание пользователей ознакомиться с содержимым антивакцинного поста и на воспринимаемую достоверность самого контента. Измерение первого показателя, например через отслеживание кликов или времени, проведенного на странице поста, позволит оценить сдерживающий эффект предупреждений. Оценка второго аспекта покажет, способствуют ли предупреждения критическому восприятию информации и снижению субъективной веры в ее правдивость, что является конечной целью подобных интервенций.

Обобщая результаты анализа, также можно сформулировать ключевые рекомендации по дизайну предупреждающих сообщений.

Необходимость пояснений. Результаты показывают, что пользователи положительно реагируют на дополнительные пояснения. Это относится не только к тексту, который опровергает конкретное заблуждение, но и к техническим пометкам, объясняющим причину появления предупреждающих сообщений. Например, пояснение, построенное по структуре «Fact — Myth — Fallacy — Fact» было охарактеризовано как «грамотно структурированное», «доступное» и «подробное». Несмотря на это, некоторые признались, что в реальности они, скорее всего, не стали бы читать объяснение, так как у них уже есть четкая позиция в отношении вакцинации. Учитывая в целом положительную реакцию и ценность пояснений для других пользователей, данные наблюдения позволяют сделать вывод, что дополнительные текстовые разъяснения являются уместным элементом, который рекомендуется включать в дизайн предупреждающих сообщений. Мы также предлагаем представлять их в виде раскрывающегося поля, чтобы не раздражать тех пользователей, которые не желают знакомиться с дополнительной информацией.

Отказ от скрытия контента. Хотя в интервью было выявлено предпочтение интерстициального типа, последующий анализ показал, что пользователям не нравится, когда социальная сеть скрывает контент от просмотра. Многие участники

DOI: 10.14515/monitoring.2025.6.2917

отметили, что интерстициальная блокировка контента может быть воспринята как цензура, нарушающая право на свободу слова. Кроме того, замечено, что закрытие текста поста предупреждающим сообщением только усиливает любопытство пользователя, а это, в представлении участников, может повышать вероятность прочтения недостоверной информации. Данные наблюдения позволяют сделать вывод, что дизайн предупреждения должен ясно показывать, что это всего лишь дополнение к основному контенту.

Указание ссылок. Для улучшения эффективности предупреждающих сообщений также рекомендуется включать ссылки на дополнительные источники. Участники отмечали, что ссылка на авторитетное мнение или научное исследование могла бы помочь им удостовериться в правильности представленных фактов. Другими словами, пользователям важно сохранять автономию в процессе взаимодействия с предупреждающим сообщением, то есть самостоятельно решать, чему доверять, а чему нет. В дизайне предупреждения также важно убедиться в том, что ссылки вызывают доверие. Некоторые пользователи сообщали, что они часто игнорируют призывы перейти на внешние источники из-за опасений стать жертвой вредоносного программного обеспечения. Вероятно, эту проблему можно решить через пометку ссылок галочкой, при наведении на которую будет отображаться информация о том, что ссылка была проверена администрацией социальной сети.

Возможность обратной связи. Наконец, результаты анализа показывают, что пользователи воспринимают возможность опротестовать предупреждение о потенциально недостоверной информации в качестве полезной функции, способной восстановить справедливость в случае ошибочного появления предупреждающего сообщения. Исходя из этого, финальной рекомендацией является внедрение механизма обратной связи в дизайн предостерегающего сообщения.

Список литературы (References)

1. Дудина В. И., Сайфулина В. О. «Почитала, еще меньше вакцинироваться захотелось»: онлайн-дискурс вакцинной нерешительности // Мониторинг общественного мнения: экономические и социальные перемены. 2023. No. 1. С. 279—298. https://doi.org/10.14515/monitoring.2023.1.2344.
Dudina V. I., Saifulina V. O. (2023) «I Read It, I Wanted to Get Vaccinated Even Less»: An Online Discourse of Vaccine Hesitancy. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 279—298. https://doi.org/10.14515/monitoring.2023.1.2344. (In Russ.)

2. Akhawe D., Felt A. P. (2013) Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In: *Proceedings of the 22nd USENIX Conference on Security*. P. 257—272. URL: https://dl.acm.org/doi/10.5555/2534766.2534789 (accessed: 17.02.2025).

3. Allen J., Watts D. J., Rand D. G. (2024) Quantifying the Impact of Misinformation and Vaccine-Skeptical Content on Facebook*. *Science*. Vol. 384. No. 6699. https://doi.org/10.1126/science.adk3451.

DOI: 10.14515/monitoring.2025.6.2917

4. Benoit S. L., Mauldin R. F. (2021) The “Anti-Vax” Movement: A Quantitative Report on Vaccine Beliefs and Knowledge Across Social Media. *BMC Public Health*. Vol. 21. No. 1. https://doi.org/10.1186/s12889-021-12114-8.

5. Clayton K., Blair S., Busam J. A., Forstner S., Glance J., Green G., Kawata A., Kovvuri A., Martin J., Morgan E., Sandhu M., Sang R., Scholz-Bright R., Welch A. T., Wolff A. G., Zhou A., & Nyhan B. (2020) Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. *Political Behavior*. Vol. 42. No. 4. P. 1073—1095. https://doi.org/10.1007/s11109-019-09533-0.

6. Epstein Z., Foppiani N., Hilgard S., Sharma S., Glassman E., Rand D. (2022) Do Explanations Increase the Effectiveness of AI—Crowd Generated Fake News Warnings? In: Budak C., Cha M., Quercia D. (eds.) *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16. P. 183—193. https://doi.org/10.1609/icwsm.v16i1.19283.

7. Fritz C. O., Morris P. E., Richler J. J. (2012) Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology: General*. Vol. 141. No. 1. P. 2—18. https://doi.org/10.1037/a0024338.

8. Gantiva C., Sotaquirá M., Marroquín M., Carné C., Parada L., & Muñoz M. A. (2019) Size Matters in the Case of Graphic Health Warnings: Evidence from Physiological Measures. *Addictive Behaviors.* Vol. 92. P. 64—68. https://doi.org/10.1016/j.addbeh.2018.12.003.

9. Guo C., Guo Z., Zheng N., Guo C. (2024) All Warnings Are Not Equal: A User-Centered Approach to Comparing General and Specific Contextual Warnings Against Misinformation. In: Bui T. (ed.) *Proceedings of the 57th Hawaii International Conference on System Sciences.* P. 2330—2339. URL: https://aisel.aisnet.org/hicss-57/dsm/critical_and_ethical_studies/3 (accessed: 12.01.2025).

10. Hassan A., Barber S. J. (2021) The Effects of Repetition Frequency on the Illusory Truth Effect. *Cognitive Research: Principles and Implications*. Vol. 6. No. 1. https://doi.org/10.1186/s41235-021-00301-5.

11. Johnson H. M., Seifert C. M. (1994) Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Vol. 20. No. 6. P. 1420—1436. https://doi.org/10.1037/0278-7393.20.6.1420.

12. Kaiser B., Wei J., Lucherini E., Lee K., Matias J. N., Mayer J. (2021) Adapting Security Warnings to Counter Online Disinformation. In: *30th USENIX Security Symposium (USENIX Security 21)*. P. 1163—1180. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/kaiser (accessed: 29.12.2024).

13. Koch T. K., Frischlich L., Lermer E. (2023) Effects of Fact-Checking Warning Labels and Social Endorsement Cues on Climate Change Fake News Credibility and Engagement on Social Media. *Journal of Applied Social Psychology*. Vol. 53. No. 6. P. 495—507. https://doi.org/10.1111/jasp.12959.

DOI: 10.14515/monitoring.2025.6.2917

14. König L. M. (2023) Debunking Nutrition Myths: An Experimental Test of the “Truth Sandwich” Text Format. *British Journal of Health Psychology*. Vol. 28. No. 4. P. 1000—1010. https://doi.org/10.1111/bjhp.12665.

15. Konstantinou L., Caraban A., Karapanos E. (2019) Combating Misinformation Through Nudging. In: Lamas D., Loizides F., Nacke L., Petrie H., Winckler M., Zaphiris P. (eds.) *Human-Computer Interaction — INTERACT 2019*. Cham: Springer. Vol. 11749. https://doi.org/10.1007/978-3-030-29390-1_51.

16. Kotz J., Giese H., König L. M. (2023) How to Debunk Misinformation? An Experimental Online Study Investigating Text Structures and Headline Formats. *British Journal of Health Psychology*. Vol. 28. No. 4. P. 1097—1112. https://doi.org/10.1111/bjhp.12670.

17. Lewandowsky S., Ecker U. K., Seifert C. M., Schwarz N., Cook J. (2012) Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*. Vol. 13. No. 3. P. 106—131. https://doi.org/10.1177/1529100612451018.

18. Martel C., Rand D. G. (2023) Misinformation Warning Labels Are Widely Effective: A Review of Warning Effects and Their Moderating Features. *Current Opinion in Psychology*. Vol. 54. Art. 101710. https://doi.org/10.1016/j.copsyc.2023.101710.

19. Mena P. (2020) Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook*. *Policy & Internet*. Vol. 12. No. 2. P. 165—183. https://doi.org/10.1002/poi3.214.

20. Nassetta J., Gross K. (2020) State Media Warning Labels Can Counteract the Effects of Foreign Misinformation. *Harvard Kennedy School Misinformation Review*. Vol. 1. https://doi.org/10.37016/mr-2020-45.

21. Pennycook G., Bear A., Collins E. T., Rand D. G. (2020) The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science*. Vol. 66. No. 11. P. 4944—4957. https://doi.org/10.1287/mnsc.2019.3478.

22. Petrov I. (2022) Anti-vaccination Movement on VK: Information Exchange and Public Concern. In: Alexandrov D. A., Boukhanovsky A. V., Chugunov A. V., Kabanov Y., Koltsova O., Musabirov I. (eds.) *Digital Transformation and Global Society*. Cham: Springer. P. 108—121. https://doi.org/10.1007/978-3-030-93715-7_8.

23. Pluviano S., Watt C., Della Sala S. (2017) Misinformation Lingers in Memory: Failure of Three Pro-vaccination Strategies. *PloS ONE*. Vol. 12. No. 7. https://doi.org/10.1371/journal.pone.0181640.

24. Porter E., Wood T. J. (2022) Political Misinformation and Factual Corrections on the Facebook* News Feed: Experimental Evidence. *The Journal of Politics*. Vol. 84. No. 3. P. 1812—1817. https://doi.org/10.1086/719271.

DOI: 10.14515/monitoring.2025.6.2917

25. Sharevski F., Alsaadi R., Jachim P., Pieroni E. (2022) Misinformation Warnings: Twitter's Soft Moderation Effects on COVID-19 Vaccine Belief Echoes. *Computers & Security*. Vol. 114. https://doi.org/10.1016/j.cose.2021.102577.

26. Silic M. (2016) Understanding Colour Impact on Warning Messages: Evidence from Us and India. In: Kaye J., & Druin A. (eds.) *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. P. 2954—2960. https://doi.org/10.1145/2851581.2892276.

27. Sullivan G. M., Artino A. R. (2013) Analyzing and Interpreting Data from Likert-Type Scales. *Journal of Graduate Medical Education*. Vol. 5. No. 4. P. 541—542. https://doi.org/10.4300/JGME-5-4-18.

28. Swire-Thompson B., DeGutis J., Lazer D. (2020) Searching for the Backfire Effect: Measurement and Design Considerations. *Journal of Applied Research in Memory and Cognition*. Vol. 9. No. 3. P. 286—299. https://doi.org/10.1016/j.jarmac.2020.06.006.

29. Tulin M., Hameleers M., de Vreese C., Opgenhaffen M., Wouters F. (2024) Beyond Belief Correction: Effects of the Truth Sandwich on Perceptions of Fact-checkers and Verification Intentions. *Journalism Practice*. P. 1—20. https://doi.org/10.1080/17512786.2024.2311311.

30. Xie J., Yamashita M., Cai Z., Xiong A. (2022) A User Study on the Feasibility of Topic-aware Misinformation Warning on Social Media. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 66. No. 1. P. 621—625. https://doi.org/10.1177/1071181322661252.