



Исследование

распространенности ненависти в эстонских социальных сетях

проведено в течение сентября 2022 года
авторы: Келли Гроссталь и Мари-Лийз Вяхи

Введение

Эстонский центр по правам человека является единственной организацией в Эстонии, которая осуществляет активный мониторинг проявлений ненависти и ведет работу по информированию об ее опасностях. Эстонский центр по правам человека является партнером Европейской комиссии в области мониторинга тенденций по разжиганию ненависти на всех платформах социальных медиа. Предусмотренная уголовным правом законодательная база, касающаяся языка вражды, в Эстонии является очень ограниченной, и уголовное преследование в связи с языком вражды иницируется достаточно редко. Поэтому бороться с ненавистью в Эстонии весьма сложно. Кроме того, активно распространяемые в последние годы фейковые новости обусловили увеличение распространенности языка вражды в социальных медиа. Поскольку язык вражды в Эстонии крайне распространен, но на государственном уровне этой проблемой почти не занимаются, целью данного исследования является стремление восполнить существующий пробел и определить распространенность и нарративы выражений ненависти, наблюдаемых в эстонских социальных медиа. Исследование было проведено в рамках проекта «Изучай. Участвуй. Меняй», цель которого – улучшить устойчивость общества к лжи и вражде, социальную сплоченность и интеграцию путем усиления роли основанных на фактах средств массовой информации, медиаграмотности, гражданской активности и сотрудничества в области образования, обучения и обменов в Эстонии. Проект финансировался Посольством Федеративной Республики Германия в Эстонии.

Анализ истории и потребностей

В общепринятой речи «язык вражды» означает оскорбительную риторику, направленную против группы лиц или отдельного лица на основании врожденных характеристик (таких как этническая принадлежность, религия, сексуальная ориентация, возраст, инвалидность или пол) и таким образом способную угрожать безопасности. Язык вражды направлен на разжигание ненависти или подстрекательство к насилию против таких групп или членов групп. Защита свободы выражения не распространяется на язык вражды, так как он представляет собой злоупотребление свободой выражения. Мы все имеем право свободно выражать себя, даже если наше мнение способно оскорбить, шокировать или обеспокоить других людей. Однако не все может быть приемлемым исходя из принципов свободы слова. Когда люди начинают публично подстрекать других к насилию, враждебности или дискриминации против определенных групп, такая риторика является не проявлением свободы слова, а языком вражды.

Это также означает, что государство может законным образом предотвращать использование языка вражды и наказывать за такие выражения. Целью языка вражды является оскорбление и унижение человека из-за того, кем он является, а не из-за того, что он сделал. Язык вражды способствовал многим катастрофам в области прав человека, начиная от нацистского холокоста и до геноцида в Руанде.

Для языка вражды всегда важен контекст, и для его оценки нужно принимать во внимание окружающую среду, говорящего и группу, являющуюся мишенью. Для определения языка вражды необходимо учитывать различные составляющие, включая контекст выражения, тон (письменная или устная форма), мишени (отдельные лица или группы лиц), а также возможные последствия или результаты. Борьба с языком вражды зачастую очень непросто, так как нередко он может являться каким-либо речевым высказыванием или одним из видов неподобающего поведения, которое может как быть наказуемым в уголовном порядке, так и считаться грубым и неприятным высказыванием, которое, тем не менее, приходится признать допустимым.

Для успешной борьбы с языком вражды, помимо распознавания и понимания данного феномена, важнейшую роль играет и юридическая сторона. Антидискриминационное законодательство Эстонии основано на статье 12 Конституции, которая запрещает дискриминацию и разжигание ненависти на почве этнической принадлежности, расовой принадлежности, религии или политических убеждений. Насилие или дискриминация также запрещены и преследуются по закону. Пенитенциарный кодекс включает положения, которые запрещают разжигание ненависти, а также нарушение равноправия в целом. Приведенное в Пенитенциарном кодексе положение о разжигании ненависти, в свою очередь, сформулировано следующим образом: *«Часть 1 статьи 151. Разжигание ненависти. Действия, публично призывающие к ненависти, насилию или дискриминации в связи с национальной или расовой принадлежностью, цветом кожи, гендерной принадлежностью, языком, происхождением, вероисповеданием, сексуальной ориентацией, политическими убеждениями, имущественным или социальным положением, если это повлекло угрозу жизни, здоровью или имуществу лица, наказываются штрафом в размере до трехсот штрафных единиц или арестом.*

Понятие разжигания ненависти или подстрекательства к насилию — юридический термин, который очень часто используется как синоним термина «язык вражды». В то же время язык вражды имеет более широкое значение по сравнению с разжиганием ненависти или подстрекательством к насилию, так как он включает в себя большее количество выражений, которые поддерживают, подстрекают, стимулируют или оправдывают ненависть, насилие и дискриминацию против лица или группы лиц по самым разным причинам.

На практике применение данного запрета ограничено, и его использовали лишь несколько раз. Например, в 2019 и 2020 годах по статье 151 не было представлено ни одного заявления (информация о 2021 годе еще не была обнародована). Проблема заключается в формулировке положения, согласно которой наказуемой является только такое разжигание ненависти, которое обуславливает непосредственную опасность для жизни, здоровья или имущества человека. Соответственно, в Эстонии возможности правоохранительных органов по привлечению использующих язык вражды лиц к ответственности достаточно ограничены.

Несмотря на то, что проявления языка вражды могут встречаться как в реальной жизни, так и в Интернете, его использование онлайн становится все более серьезной проблемой, способной оказать серьезное негативное влияние на психическое здоровье человека, а также его вовлеченность в онлайн-жизнь и благополучие в целом. Кроме того, в самых экстремальных случаях язык вражды может привести к преследованию и насилию уже в реальной жизни. Последние десятилетия мы наблюдали стремительный рост использования Интернета и платформ социальных медиа. Однако это сопровождалось и ростом использования языка вражды в Интернете, в связи с чем государственные институты и организации Европейского Союза создали целый ряд политических мер, направленных на разрешение данного феномена. Можно утверждать, что охватившая весь мир пандемия Covid-19 привела к большему времяпровождению в Интернете, что, в свою очередь, вызвало и значительную активизацию использования языка вражды онлайн. Недостаточность систематического модерирования со стороны ИТ-компаний привела к тому, что Интернет становится пространством, способствующим усилению влияния экстремистов и распространению языка вражды. Несмотря на то, что Европейская комиссия и общественные организации предприняли различные меры по противодействию языку вражды, необходимость осуществления мониторинга языка вражды и незаконного контента продолжает оставаться актуальной.

Методология мониторинга социальных медиа

В течение одного месяца были проведены две сессии мониторинга для исследования публикаций в социальных медиа, размещенных как русскоязычными, так и эстоноязычными жителями Эстонии. В результате была обнаружена 51 содержащая язык вражды публикация на эстонском языке и 52 такие публикации на русском. В данном исследовании сравнивается содержание публикаций, размещенных на эстонском и на русском языке, и идентифицируются существенные различия в нарративах. В случае каждой публикации отмечалась платформа социальных медиа и три существенных аспекта (нарратива)..

Целью первого вопроса относительно нарратива было определить защищаемую группу, которая являлась мишенью публикации. Чтобы получить больше сравнительных факторов для анализа, были выбраны две защищаемые группы. Первая группа была образована исходя из гендерной и сексуальной ориентации, а вторая – исходя из национальности, этнической и расовой принадлежности. Данные две группы были выбраны в связи с тем, что в последние годы в эстонском обществе именно эти две категории групп регулярно и систематически являлись объектом языка вражды (в соответствии с результатами предыдущих мониторинговых сессий, проведенных центром).

Целью второго вопроса относительно нарратива было определить тип ненависти, выраженной в таких публикациях. Было определено, содержался ли в разжигающей ненависть публикации прямой призыв к действию, разжигание ненависти, заявления о неполноценности, выражение презрения, отвращения или пренебрежения, ругань и призывы к изгнанию или сегрегации. Целью третьего вопроса относительно нарратива было определить, использовался ли в публикации прямой язык или метафора. Во время мониторинга были также определены и отмечены тенденции и модели выражения ненависти (например, конкретные используемые слова), обнаруженные в социальных медиа.

О каждой размещенной в социальных медиа публикации, содержащей язык вражды, платформе было отправлено сообщение о недопустимой публикации. Затем организаторы мониторинга фиксировали, насколько быстро платформы удаляли такие содержавшие язык вражды публикации, делали ли они это вообще, и влиял ли на их реакцию язык публикации. Для проведения мониторинга были выбраны четыре популярные платформы социальных медиа – Facebook, Instagram, TikTok, Twitter. В будущем в проведение аналогичного мониторинга языка ненависти, используемого в Эстонии на русском языке, также может быть включена платформа Telegram, однако в рамках данного сравнительного исследования рассматривались только четыре вышеупомянутые платформы.

Результаты исследования

В целом, организаторы мониторинга отметили, что количество содержащего ненависть и неприязнь контента было большим. Тем не менее, не весь такой контент мог считаться языком вражды или разжиганием ненависти, нарушавшим стандарты сообщества или правила поведения платформ социальных медиа (а также местное законодательство). В случае обоих языков было обнаружено, что люди научились распространять ненависть, обходя основанные на машинном обучении модели определения ненависти в социальных медиа, например посредством использования

метафор и скрытых угроз. Кроме того, организаторы мониторинга отметили, что люди предупреждали друг друга о возможности отправки сообщения о недопустимой публикации и последующей блокировке пользователя. Во время мониторинга были определены социальные медиа, платформы, группы и страницы, которые содержали контент, выражавший ненависть более агрессивно. Мониторинг ненависти, выраженной на русском языке, проводился в русскоязычных группах ЛГБТ+ в Facebook. Так как в настоящее время в СМИ публикуется меньше новостей на тему ЛГБТ+, некоторые из комментариев, о которых были отправлены сообщения как о недопустимых, были опубликованы несколько лет назад.

Язык ненависти, использовавшийся на эстонском языке, был обнаружен на таких размещенных в Facebook страницах, как Delfi, Postimees, Objektiiv, Eestinen, Uued Uudised, а также под видео из TikTok, например в комментариях к видео, размещенным эстонскими блогерами, цвет кожи которых не является белым или которые являются трансгендерами. В отличие от публикаций известных людей, на страницах таких СМИ, как Delfi и Postimees, было обнаружено множество комментариев, содержащих выражение ненависти – возможно, в связи с тем, что они привлекают много общественного внимания, и для мониторинга, удаления такого большого количества комментариев, а также сообщения о них просто не хватает ресурсов. В то же время создалось впечатление, что пользователи, являющиеся физическими лицами, следили за чистотой своих учетных записей, просматривая комментарии под своими публикациями или даже и вовсе отключая в них функцию размещения комментариев. Было также отмечено, что в Эстонии более молодые люди используют для самовыражения в большей степени TikTok, тогда как старшее поколение предпочитает Facebook. Комментарии в TikTok и Instagram, как правило, были более короткими и примитивными по сравнению с комментариями в Facebook или Twitter.

Нарративы языка вражды, определенные во время мониторинга социальных медиа

В течение всей мониторинговой сессии было обнаружено 103 случая использования языка вражды. Среди выбранных крупных платформ социальных медиа самым популярным инструментом для распространения содержащих ненависть сообщений был Facebook – согласно результатам мониторинга, в 87 случаях такой контент был размещен в Facebook, в 12 случаях в Tiktok, в 2 случаях в Instagram и в 2 случаях в Twitter.

Большинство таких сообщений (т. е. 80 комментариев) были нацелены против людей в связи с их гендерной и сексуальной ориентацией (главным образом, в связи с

последней), тогда как остальные сообщения (т. е. 23 комментария) были нацелены против групп исходя из национальности, этнической принадлежности и расовой принадлежности. В случае обеих групп практически во всех комментариях использовались ругательства. В первой группе мишенью в связи с сексуальной ориентацией часто становились группы людей. При использовании обоих языков не являющиеся гетеросексуалами люди часто сравнивались с педофилами, зоофилами или обезьянами, или же утверждалось, что они являются переносчиками болезней. В эстоноязычных комментариях, размещенных против групп лиц в связи с их сексуальной ориентацией, самым распространенным словом было «pede», соответствующее англоязычному «fag». Слово «pede» использовалось в девяти комментариях, о которых были отправлены сообщения как о недопустимых. На русском языке было зафиксировано больше комментариев, в которых прямо угрожающий язык использовался против людей, не являющихся гетеросексуалами (например, призывы «сжигать их»), по сравнению с эстоноязычным контентом, в случае которого в основном выражались послы о неполноценности, или демонстрировалось презрение, отвращение или пренебрежение. В нескольких случаях были отмечены содержащие ненависть комментарии против трансгендеров, которых называли «оно», или в отношении которых заявлялось, что «трансгендерная женщина – это не женщина».

В комментариях, нацеленных против людей в связи с их национальностью, этнической принадлежностью или расовой принадлежностью, самыми частыми мишенями выражающих ненависть комментариев были чернокожие люди (17 таких комментариев). Слово «neeger», соответствующее англоязычному «nigger», использовалось четыре раза. В комментариях против людей африканского происхождения люди призывали к колонизации, рабству или сегрегации. Обнаруженные в ходе мониторинга послания заключались в том, что люди африканского происхождения ленивы, «если только они не находятся под властью белых», или что люди африканского происхождения должны «убраться домой», что мы должны отправить их «обратно в Африку», или что «им тут не место». Было зафиксировано четыре содержащих ненависть комментария против русских, один против украинцев и один против арабов.

Большинство сообщений включали элементы выражения презрения, отвращения или пренебрежения, однако были обнаружены и сообщения, содержавшие призывы к изгнанию или сегрегации, заявления о неполноценности, разжигание ненависти, а также прямые призывы к действиям (включая подстрекательство к насилию). При этом в некоторых комментариях использовался язык, в котором заявления о неполноценности сочетались с выражением презрения, отвращения или пренебрежения. Было отмечено, что в 93 таких случаях группа людей прямо являлась мишенью языка ненависти, тогда как в остальных 10 комментариях для выражения содержащего ненависть послания использовались метафоры. Среди содержащих язык вражды посланий наиболее частыми были призывы «сжечь себя».

В русскоязычном контенте в сообщениях использовался исключительно прямой язык. В части эстонязычного контента 41 сообщение из 51 было написано с использованием прямого языка, а в случае остальных 10 сообщений использовались метафоры. В отличие от содержащих язык вражды и направленных против сообщества ЛГБТ+ комментариев на эстонском языке, в соответствующих русскоязычных комментариях содержались призывы к травле детей ЛГБТ+. В эстонском языке часто использовались такие слова, как «pede» или «peeget». В русском языке также использовалось множество похожих слов, направленных против сообщества ЛГБТ+. Так, использовавшимися против защищаемых групп людей в связи с их сексуальной ориентацией словами были «голубой», «петушара», «пидар», «педики», «пидарюги», «гомосек». Против защищаемых групп людей в связи с их гендерной ориентацией использовались такие слова, как «извращенец», «мерзость», «твари», «гнойные», «педофилы».

Реакция социальных медиа на сообщения о недопустимости

Организаторы мониторинга сообщали соответствующим службам платформ социальных медиа о размещенном на платформах контенте, который содержал язык вражды или разжигание ненависти. Сообщения включали информацию о содержании публикации и визуальное доказательство, с указанием конкретных правил сообщества, которые нарушил такой контент. Из 103 публикаций, о которых были отправлены сообщения, 38 (26 в Facebook, 12 в TikTok) были удалены в первый же день – 28 из них были размещены на эстонском, и 10 на русском языке. 30 публикаций Facebook в первый день не посчитал нарушающими стандарты сообщества и решил не удалять их. На второй день Facebook решил не удалять еще одну публикацию. На второй день Instagram удалил одну из публикаций, о которых ему поступило сообщение, а на третий день – еще одну. На третий день Facebook удалил еще две публикации и решил, что четыре другие публикации не нарушают стандарты сообщества и, соответственно, не будут удалены. Через неделю после отправки сообщений о недопустимости никаких дополнительных публикаций удалено не было. Таким образом, из 103 публикаций, относительно которых были отправлены сообщения о недопустимости, были удалены 42 публикации. Это означает, что платформы социальных медиа удалили только 40,78% публикаций, о недопустимости которых они получили сообщения.

В течение периода проведения мониторинга (1 неделя) Twitter не предпринял никаких действий в отношении такого контента. Instagram удалил обе публикации (на второй и на третий день). TikTok удалил все публикации, относительно которых он получил сообщение о недопустимости (12) уже в первый день. Facebook удалил

29 из 87 публикаций, о которых он получил сообщения (18 на эстонском и 11 на русском языке). Это означает, что Facebook предпринял меры только в отношении 33% публикаций, о недопустимости которых ему было сообщено. Кроме того, организаторы мониторинга были удивлены тем, что Facebook не удалил некоторый контент, который явно представлял собой язык вражды и был написан с использованием прямого языка, при том, что платформа удалила некоторые комментарии, которые выражали ненависть в меньшей степени и менее прямолинейно (хотя и нарушали стандарты сообщества Facebook). Таким образом, организаторы пришли к выводу, что в отношении языка вражды система мониторинга Facebook является неясной и несистематичной. При этом на значительное количество сообщений о недопустимых комментариях (20) работники службы поддержки Facebook не отреагировали совсем. Значительной разницы между модерированием контента на эстонском и русском языках со стороны Facebook замечено не было.

Рекомендация

- Данное исследование подтвердило результаты предыдущих мониторингов, проведенных центром, согласно которым использование языка вражды в Эстонии продолжает оставаться проблемой, особенно в интернет-пространстве. Таким образом, существует необходимость в срочном порядке подготовить национальный план действий по борьбе с языком вражды, предусматривающий сотрудничество между деятелями разных сфер.
- Необходимо больше общественных дискуссий относительно понятия языка вражды и свободы слова. В результате недопонимания и дезинформации в обществе распространено множество ошибочных представлений о данном вопросе. Государство должно инициировать или финансировать программы по улучшению информированности общества о последствиях языка вражды для подверженных ему групп и общества в целом.
- Так как большинство выражающих ненависть публикаций, о недопустимости которых были отправлены сообщения, были нацелены против людей в связи с их гендерной или сексуальной ориентацией (в основном против последнего), дети и подростки должны получать в школе образование относительно гендерного разнообразия, включая вопросы, связанные с ЛГБТ-сообществом. Психологические и социологические исследования

показывают, что гетеронормативность, неприятие гомосексуализма и негативное отношение к ЛГБТ-сообществу связаны с более низким уровнем образования и интеллекта. Для учащихся, демонстрирующих гендерно-неконформное поведение, школы могут создать безопасный климат, если им удастся уменьшить гомофобную и трансфобную дискриминацию.

- В связи с тем, что некоторые политические партии участвуют в распространении содержащих ненависть и дискриминационных сообщений, важно стимулировать партии принимать кодексы этики и внедрять инструменты, позволяющие реагировать на нарушение политиками таких кодексов.
- В настоящее время наблюдается недостаток информированности и нежелание решать проблему языка вражды, что, в свою очередь, способствует сохранению обстановки безнаказанности для злоупотребляющих этим положением лиц. Таким образом, образование по правам человека в данной области, а также информационная и разъяснительная деятельность являются крайне необходимыми.
- В Эстонии регулирующие язык вражды законы являются слишком мягкими, и соответствующие положения Пенитенциарного кодекса в текущей формулировке бессильны против языка вражды, так как они требуют наличия прямой опасности в отношении жизни или благополучия лица вследствие таких слов. В Пенитенциарный кодекс необходимо внести изменения, позволяющие государству реагировать на разжигание ненависти и призывы к насилию.
- Для выяснения причин, степени распространенности и влияния языка вражды в Эстонии необходимо провести тщательные исследования. Чтобы обеспечить более эффективную борьбу с языком вражды, нацеленного против языка и культуры, необходимо продолжать более масштабные и регулярные исследования настроений и нарративов, распространяемых в социальных медиа Эстонии как на эстонском, так и на русском языке.