

Как бороться с ненавистью в социальных сетях? | Comment réagir aux discours de haine sur les médias sociaux?

Author: Заррина Салимова, [Цюрих](#), 10.12.2021.



Фото: Dan Edge, Unsplash

Цюрихские исследователи протестировали три различные стратегии ответной реакции на агрессивные высказывания в Twitter. Наиболее эффективными оказались комментарии, основанные на эмпатии.

Des chercheurs zurichois ont testé trois stratégies différentes pour répondre à des commentaires agressifs sur Twitter. Les commentaires les plus efficaces étaient ceux basés sur l'empathie.

Comment réagir aux discours de haine sur les médias sociaux?

Язык вражды (англ.: Hate Speech) в Интернете стал проблемой мирового масштаба. В социальных сетях нападают на сексуальные меньшинства, запугивают членов определенных религиозных общин и этнических групп. Изменения к худшему в поведении пользователей заметны, в частности, по комментариям к нашим публикациям в Facebook. Под [анонсом](#) интервью со знаменитой российской писательницей Людмилой Улицкой появились отвратительные антисемитские высказывания, а каждый обзор ситуации по коронавирусу заканчивается перебранкой, причем в результате во всех грехах виноватыми оказываются журналисты: мол, если бы не СМИ, которые вступили в «сговор с фармакомпаниями», то ни вируса, ни эпидемии бы не было. Да уж, как говорил Шурик, и часовню тоже мы развалили.

На многих платформах есть фильтры, которые автоматически блокируют подобные комментарии, но этих мер зачастую недостаточно. Facebook, например, по собственной оценке (или согласно внутренним документам, просочившимся в СМИ несколько месяцев назад), в состоянии удалить не более 5% высказываний ненависти. Кроме того, автоматические фильтры являются неточными и, как правило, противоречат свободе выражения мнений. Таким образом, ответственность за противодействие языку ненависти ложится на модераторов социальных сетей.

Как же реагировать на неадекватные высказывания в сети? Чтобы выяснить ответ на этот вопрос исследователи Цюрихской федеральной политехнической школы (ETHZ) и Цюрихского университета (UZH) провели [эксперимент](#). Используя методы машинного обучения, они выявили 1350 англоязычных пользователей Twitter, которые ранее публиковали расистские или ксенофобские твиты. Ученые и студенты под руководством профессора политического анализа ETHZ Доминика Хангартнера отвечали этим пользователям, используя различные стратегии. Первой группе они отвечали сообщением, призванным вызвать сочувствие, например, «Афроамериканцам очень больно, когда люди используют такие выражения». Второй группе они выносили предупреждение о возможных социальных последствиях твита, а третьей отправляли шуточный мем. Так, один из мемов изображал птицу, сжимающую клюв другой птицы с фразой «Пожалуйста, прекратите писать в Twitter». Четвертая же группа была контрольной.

Выяснилось, что только комментарии, основанные на эмпатии, способны снизить уровень языка ненависти – хотя и в относительно небольшой степени. Эта группа отправила меньше ксенофобных твитов в последующие четыре недели, чем контрольная группа, и чаще удаляла первоначальный ксенофобный твит. Исследователи также фиксировали, реагировали ли пользователи на обращенные к ним ответы и каким образом. Согласно результатам, около 15% из них ответили на сообщения – большинство сделали это достойно, а некоторые даже извинились, причем лишь незначительная часть из них использовала в своем ответе язык ненависти.

Таким образом, ответы, базирующиеся на сочувствии, оказались более эффективным средством против вражды в Интернете, чем противодействие ненависти с помощью

предупреждений или юмора. В то время как даже намек на то, что семья, друзья и коллеги комментирующего также видят его расистские сообщения, не возымел действия. Другими словами, цюрихским ученым удалось доказать тезис, который выдвигал еще Лев Толстой. Напомним, что великий русский писатель был убежден в том, что человек должен перестать отвечать злом на зло и не должен противиться злу насилием, так как в этом случае он сам становится объектом зла, насилия и ненависти.

Авторы исследования, результаты которого были опубликованы в научном журнале *Proceedings of the National Academy of Sciences*, признают, что они, конечно, не нашли универсального средства против языка ненависти в Интернете, но обнаружили важные подсказки о том, какие стратегии могут сработать, а какие нет. В будущем еще предстоит выяснить, все ли основанные на эмпатии ответы имеют одинаковый эффект. Например, комментирующих можно попросить поставить себя на место жертвы или спросить их, что бы они чувствовали, если бы люди говорили о них таким же образом.

Статьи по теме

[За что ненавидят Грету Тунберг?](#)

[Кому принадлежит наша информация из социальных сетей?](#)

[Facebook затягивает швейцарцев](#)

[Facebook недостаточно сотрудничает со швейцарскими властями?](#)

[Роботы сгенерировали половину сообщений в Twitter по теме No Billag](#)

[Twitter-карта Швейцарии и остального мира](#)

Source URL:

<http://www.nashagazeta.ch/news/la-vie-en-suisse/kak-borotsya-s-nenavistyu-v-socialnyh-setyah>