



DOI: 10.22363/2313-1438-2022-24-3-517-529

Научная статья / Research article

Политические последствия цифровизации риторики ненависти в эпоху постправды: влияние на эмоциональные режимы в ходе цифровых конфликтов

Ю.Ю. Колотаев  

*Санкт-Петербургский государственный университет, Санкт-Петербург,
Российская Федерация*

 yury.kolotaev@mail.ru

Аннотация. Повсеместное проникновение цифровой реальности в общественную жизнь ведет к видоизменению старых явлений. Некоторые из этих изменений имеют сильные негативные последствия в отдельных сферах, в частности в политике. Цифровая риторика ненависти (hate speech) является одним из наиболее показательных примеров влияния цифровизации на политические процессы. Проявление ненависти в сети стало серьезным вызовом для политических систем по всему миру. Для противодействия новой угрозе требуется ее теоретическое и практическое осмысление. Целью данной работы является выявление социальных механизмов, делающих риторику ненависти инструментом в информационных кампаниях. Для этого в рамках работы рассматриваются дискурсивные и эмоциональные аспекты, связанные с публичной манифестацией ненависти. Теоретической основой выступает теория «эмоциональных режимов» и концепция «дискурсов правды», выражающие взаимовлияние субъективности и общественного дискурса. Сопоставление теоретических рамок с практическими аспектами риторики ненависти демонстрирует, что цифровые платформы и социальные сети формируют среду, ускоряющую и облегчающую проникновение ненависти в общественное пространство. Результатом этого процесса становится принятие различных форм ненависти в качестве новой социальной нормы, что приводит к таким явлениям, как притеснение, унижение или даже физическое преследование. Современные технические реалии позволяют инструментализировать риторику ненависти в целях манипуляций. Подобные манипуляции приводят к трем основным сценариям воздействия на общество: 1) широкое воздействие с низкой персонализацией; 2) точечное воздействие с высокой персонализацией; 3) перспективный сценарий децентрализованного широкого точечного воздействия с применением искусственного интеллекта. Ключевой задачей в контексте противодействия каждому из сценариев является сочетание долгосрочных и специализированных мер.

© Колотаев Ю.Ю., 2022



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Ключевые слова: риторика ненависти, язык вражды, постправда, эмоциональный режим, онлайн-платформы

Для цитирования: Колотаев Ю.Ю. Политические последствия цифровизации риторики ненависти в эпоху постправды: влияние на эмоциональные режимы в ходе цифровых конфликтов // Вестник Российского университета дружбы народов. Серия: Политология. 2022. Т. 24. № 3. С. 517–529. <https://doi.org/10.22363/2313-1438-2022-24-3-517-529>

Political Implications of Hate Speech Digitalization in a Post-Truth Era: Impact on Emotional Regimes in Digital Conflicts

Yury Y. Kolotaev  

Saint Petersburg State University, Saint Petersburg, Russian Federation

 yury.kolotaev@mail.ru

Abstract. The digitalization of public life modifies old phenomena. Some of these changes prove to be detrimental to certain public spheres, including politics. Digital hate speech is one of the most obvious examples of digitalization’s impact on political processes. The manifestation of hatred online became an urgent challenge to political systems globally. In this respect, a theoretical and practical elaboration is necessary to counter the new threat. This study aims to identify the social mechanisms that make hate speech an instrument in information campaigns. The article illustrates the discursive and emotional aspects of the public manifestation of hatred. The theoretical basis of this work is the theory of “emotional regimes” and the concept of “regimes of truth”, which express the mutual influence of subjectivity and public discourse. Comparing the theoretical framework with the practical aspects of hate speech demonstrates that digital platforms and social networks form an environment that accelerates and facilitates the dissemination of hatred in the public space. As a result, various forms of hatred are accepted as a new social norm, which leads to such phenomena as harassment, humiliation, or even physical persecution. The modern technical reality allows to instrumentalize hate speech for manipulation, which results in three main scenarios for social impact: 1) large impact with low personalization; 2) targeted impact with high personalization; 3) broad and decentralized targeted impact using artificial intelligence. While countering each of these scenarios, a key challenge is to combine long-term and specialized measures.

Keywords: hate speech, post-truth, emotional regimes, online platforms

For citation: Kolotaev, Y.Y. (2022). Political implications of hate speech digitalization in a post-truth era: Impact on emotional regimes in digital conflicts. *RUDN Journal of Political Science*, 24(3), 517–529. (In Russian). <https://doi.org/10.22363/2313-1438-2022-24-3-517-529>

Введение

Современные цифровые медиа, служащие одновременно и информационным каналом, и средством коммуникации, создали среду, в которой классические информационные угрозы приобрели новую форму и содержание. За последнее десятилетие активизировались проблемы, связанные с распространением ложной или вредоносной информации, способной нанести существенный социальный урон. В научном и публицистическом пространстве это явление получило

название *политика* или *эпоха постправды* [Попова и др. 2018; Kalpokas 2019; Cosentino 2020; Vufacchi 2021]. Ключевыми характеристиками, свойственными этим условиям, становятся децентрализация источников информации, беспрецедентный рост цифровых данных и персонализированного контента при явной социальной поляризации, приводящей к изоляции различных дискурсов и снижению значимости верификации в социальном пространстве, уступающей место эмоциям [d'Ancona 2017]. Иначе говоря, «постправда» — это ситуация, в которой фактические данные перестают иметь универсальное значение, так как информационный поток любого пользователя ведет к созданию субъективного, эмоционального потока информации, соответствующего мировоззрению пользователя.

Данные условия становятся удобной средой для манипулирования или оказания давления на общество в политических целях. На сегодняшний день эта проблема получает все большее рассмотрение в работах отечественных [Субочев 2019; Мартыанов и др. 2019; Золян С.Т. и др. 2021] и зарубежных исследователей [d'Ancona 2017; Ali, Zain-ul-abdin 2021]. В особенности активно развивается изучение проблемы цифровой дезинформации или фейковых новостей [Morgan 2018; Золян и др. 2021; Karantai et al. 2021]. Вместе с тем информационные манипуляции в форме распространения заведомо ложных сведений являются не единственной угрозой, проистекающей от проникновения цифровых медиа в социальную реальность. Одной из наиболее актуальных проблем становится «риторика ненависти» или «язык вражды» (от англ. hate speech) [Хлопотунов 2020; Sellars 2016; Castaño-Pulgarín et al. 2021], т.е. публичное проявление ненависти к какому-либо индивиду или сообществу через оскорбительные или уничижительные высказывания. В рамках Стратегии и плана действий ООН по борьбе с риторикой ненависти обращается внимание на то, что поводом для нее могут выступать любые коммуникативные акты, нацеленные на дискриминацию по каким-либо групповым или индивидуальным признакам¹. Поводом для риторики ненависти могут быть политические, этнические (и т.д.) различия [Matamoros-Fernández, Farkas 2021]. Подобные враждебные акты и высказывания в сети расширяют представление о вредоносном контенте в сети за рамки простой лжи и фейков.

Риторика ненависти сама по себе не является уникальным явлением, так как присуща общественной коммуникации в условиях конфликта социальных групп и личностей. Однако цифровая реальность повысила скорость ее распространения и обострила проблему посредством новых инструментов и платформ, способных превратить язык ненависти в инструмент стратегических коммуникаций или информационного воздействия. В данной статье раскрываются ключевые аспекты инструментализации языка вражды в рамках цифровых конфликтов. Целью является установление теоретических и практических предпосылок, по-

¹ ООН. Стратегия и план действий Организации Объединенных Наций по борьбе с ненавистнической риторикой. 2019. URL: https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_RU.pdf (дата обращения: 06.06.2022).

звolyающих использовать риторику ненависти в информационных кампаниях. Теоретической основой служит теория «эмоциональных режимов» [Reddy 2001; Wahl-Jorgensen 2018] в сочетании с концепцией «дискурсов правды» Мишеля Фуко [Foucault 2017]. В статье рассматривается взаимное влияние социального конструирования правды на общественное восприятие «ненависти».

Эмоциональные режимы, ненависть и (пост-)правда как взаимосвязанные дискурсивные категории

Формирование современной цифровой «эпохи постправды» имеет глубокие социальные корни, уходящие в область общественного конструирования категории «правды» и ее влияния на эмоциональный отклик людей на факты [Попова и др. 2018]. Связь между правдой и эмоциями обусловлена двойственностью каждой из категорий. С одной стороны, в общественном пространстве правду, как понятие близкое к фактам, принято противопоставлять эмоциям, как явлениям психологического характера [Sullivan 2013]. Вместе с тем существуют и альтернативные воззрения, характеризующие эмоции (в частности и ненависть) как проявление «заученных когнитивных паттернов» [McGrath 2017; Süselbeck 2019: 282]. Такой подход, получивший название «история эмоций» [Reddy 2001; McGrath 2017], позволяет трактовать эмоции шире, учитывая культурное и социальное влияние.

В контексте такого видения, в общественных науках появилось понятие «эмоциональных режимов», выражающих доминирующие в обществе модели восприятия эмоций [Reddy 2001]. Эмоциональные режимы формируются через нормы, обычаи и политические институты, что позволяет говорить об их «управляемости» и интерпретируемости [Colwell 2016: 7]. Особенно характерно проявление эмоциональных режимов в контексте «политических эмоций», таких как ненависть и ярость [Lytan 2004: 133]. Их политическая составляющая проистекает от того, что они обладают особенно выраженной способностью мотивировать социальные общности на активное проявление своей политической воли. При этом, воспринимая эмоции как когнитивное явление, следует отметить, что важное влияние на восприятие ненависти оказывает окружающая людей реальность и «правда» как характеристика отношения и восприятия людьми реальности. Те явления, которые воспринимаются людьми как приемлемые, в контексте эмоционального режима могут трактоваться как правдивые или «оправдывающие». Эти обстоятельства во многом усиливаются в эпоху постправды, когда на передний план выходит эмоциональная связь с фактами. Однако существует и обратная взаимосвязь между правдой и влиянием на эмоциональный режим.

С точки зрения постструктурализма этот аспект раскрывается посредством обращения к категории «дискурсов правды», разработанной М. Фуко [Foucault 2017]. В его понимании существование дискурса, претендующего обозначить правду, не означает, что существует прямая связь между дискурсом и той реальностью, о которой он говорит [Foucault 2017: 221]. Фуко проблематизирует параллельное существование реальности и правды [Foucault 2017: 237]. Для него реальность мира не является правдой самой для себя, так как дискурс-

сы правды не только документируют, что является частью реальности, а сами представляют собой фрагмент реальности, которая не нуждается в их появлении [Prozorov 2019: 21]. Правда, таким образом, является сущностью не обязательной для реальности, но дополняющей ее в случае своего появления. То обстоятельство, что эмоции во многом определяют наше понимание «реальности», указывает на то, что дискурс правды в равной степени зависит от действующего эмоционального режима, как и сам режим от дискурса правды.

Взаимное дополнение представлений о дискурсах правды и эмоциональных режимах позволяет создать комплексное представление, объясняющее предпосылки для роста угрозы от дезинформации и риторики ненависти. Если сама «контролируемость» и эмоций, и правды в социально-политическом плане существует с момента формирования любых сообществ, то ускорение через цифровизацию динамики изменения коллективных сентиментов, равно как и сила новых рычагов воздействия, меняет характер этих процессов.

Ненависть как политическая эмоция и ее цифровая инструментализация

Ключевым свойством языка вражды, как политической категории, становится ее связь с одним из наиболее активных стимулов в рамках общественного дискурса — ненавистью. Распространение в современных условиях политкорректности и снижение социальной приемлемости агрессии вызвали особое порицание ненависти из-за увеличения внимания к чувствам и стремлению не допускать их оскорбления. Однако повышенное внимание к темам эмоционального или психологического насилия не снизило значимость ненависти как политического триггера. В отдельных случаях ненависть стала культивироваться как обратная реакция на политкорректность [Wilber 2017].

Важно отметить, что ненависть позволяет направлять недовольство в конкретное русло, превращая его в совокупность действий, варьирующихся от уничижения объекта ненависти, до прямого насилия [Chetty, Alathur 2018]. В этой связи ненависть имеет тесную связь с насильственными правонарушениями и даже экстремизмом. Действия различных экстремистских групп часто сопровождаются разжиганием ненависти в сети [Olteanu et al. 2018]. При этом, даже при отмечаемом глобальном снижении уровня насилия [Pinker 2011], цифровое насилие или насилие, провоцируемое языком вражды, остается актуальным, особенно в условиях роста количества ненавистнического контента.

Вопрос культивации ненависти связан с формируемыми в политическом пространстве представлениями о приемлемости тех или иных форм языка вражды. Если один контекст предусматривает жесткое порицание ненависти, то другой смещает эмоциональный режим в сторону более открытого выражения неприязни, что может спровоцировать серьезные последствия. К примеру, в конфликтной ситуации общество предрасположено к открытой неприязни по отношению к оппоненту. В этом случае дискурс правды, находящийся под давлением эмоционального режима [Reddy 2001; Colwell 2016], зависит

от действий политических и социальных акторов. И если раньше такими акторами выступали государства (со своей правовой системой) и СМИ (с собственной редакторской этикой), то теперь большую роль начинают играть *цифровые платформы*, вмешивающиеся в процесс информационной модерации. Вместе с тем суть вовлечения платформ состоит не только в модерации, но и в самом создании универсальных площадок для распространения информации.

Открытость цифровых платформ дает доступ к информационному воздействию широкому кругу сторон. Пользователи и различные политические группы имеют открытые рычаги влияния на эмоциональный режим сообщества через распространение различного, зачастую эмоционально окрашенного контента. При этом, при всем разнообразии эмоциональных сообщений в интернете, именно риторика ненависти вызывает наибольшие опасения. Об этом свидетельствует наличие специальных программ, мер или законов на глобальном, региональном и национальном уровнях.

Каждому из уровней свойствен собственный акцент в предпринимаемых мерах. Так, на глобальном уровне особое место занимают адаптивные меры. Их основным свойством выступают гибкие механизмы взаимодействия общества, государства и бизнеса. Примерами глобальных инициатив выступают Стратегия ООН², а также рекомендации ЮНЕСКО [Gagliardone et al. 2015], определяющие этические кодексы, исследования риторики ненависти, а также развитие цифровой грамотности и критического мышления в качестве ключевых мер. В силу их рекомендательного характера саморегуляция выступает основной моделью противодействия. В сравнении с глобальными мерами региональные инициативы смещают основной акцент в пользу совместного регулирования. К примеру, Европейский союз³ стремится расширить сферу права ЕС для обеспечения юридических обязательств со стороны цифровых платформ. Однако до сих пор основные инициативы имеют необязательный характер.

Нехватка регулятивного потенциала международных организаций компенсируется нормотворчеством на национальном уровне. Происходит это через создание новых законов или применение старых норм к цифровому пространству. При этом государства избирают различные механизмы ответственности. К примеру, в рамках немецкого законодательства ответственность за язык вражды и, соответственно, за модерацию лежит на цифровых платформах. В других странах, в частности и в РФ, помимо платформ ответственность несут и сами пользователи. Вместе с тем возложение ответственности на различных субъектов за язык вражды вызывает немало опасений касательно ограничений свободы слова [Gagliardone et al. 2015; Pitruzzella, Pollicino 2020], что, в зависимости

² ООН. Стратегия и план действий Организации Объединенных Наций по борьбе с ненавистнической риторикой. 2019. URL: https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_RU.pdf (дата обращения: 06.06.2022).

³ Proposal for a *Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final*. Brussels, 2020.

от юридической практики и культурных особенностей, приводит к различным правоприменительным практикам.

Артикуляция риторики ненависти на глобальном и национальном уровнях как принципиальной цифровой угрозы совпала с ростом социальных сетей. В современном мире множится количество примеров, когда через цифровой язык вражды происходит не только сетевая травля различных меньшинств или социальных групп, но и их физическое преследование. Наиболее показательным в этой связи стал геноцид в Мьянме, где ключевую роль в распространении ненависти сыграли именно социальные сети [Siddiquee 2020]. Аналогичная ситуация складывалась в период пандемии COVID-19, когда сторонники и противники вакцинации в разных странах мира подвергали друг друга сетевой травле. В актуальных же реалиях наблюдается начало аналогичного процесса в отношении российского сегмента интернета после нового витка украинского кризиса.

С технической точки зрения, распространение риторики ненависти в ходе социального конфликта может происходить по нескольким каналам и с различными целями. Каналами чаще всего выступают информационные ленты цифровых платформ и соцсетей, распространяющие эмоционально привлекательные заголовки. Такие каналы могут пользоваться системой таргетинга самой социальной сети либо сбором открытых данных о пользователях. Цифровая революция позволила существенно увеличить масштаб и качество самих кампаний. При этом для персонализации риторики ненависти активно привлекаются и передовые технологии, в частности искусственный интеллект, способный на основе данных формировать систему определения предпочтений пользователей. Такие тенденции делают цифровое пространство все более привлекательным для манипуляций.

Следует учитывать, что риторика ненависти не всегда основана на внешних манипуляциях. Пользователи могут самостоятельно, без внешнего вмешательства, распространять контент, наносящий социальный ущерб, личный урон или вызывающий общественную поляризацию. Архитектура современных платформ, персонализирующих информационный поток пользователей, делает их информационное поле более изолированным, снижая понимание и терпимость к членам других информационных систем или пузырей [Flaxman, Goel, Rao 2016]. В этой связи при разработке мер по противодействию цифровой риторике ненависти важно учитывать, что сама угроза имеет различные источники и масштабы, но при этом сохраняет в каждом случае значительный потенциал к социальной поляризации. Особое значение в этом вопросе занимают цифровые платформы, являющиеся одновременно и инструментом в руках распространителей языка вражды, и актором, способным ограничить ее распространение.

Сценарный анализ социального воздействия риторики ненависти и перспективы противодействия

Наибольшую опасность в контексте эмоциональных режимов риторика ненависти представляет в форме манипулятивного инструмента в рамках цифрового конфликта (т.е. намеренное использование языка вражды для трансформации

общественных настроений). Логику манипулятивной атаки позволяют объяснить три базовых сценария, основывающиеся на двух переменных: масштаб кампании и степень персонализации.

Первый потенциальный сценарий включает в себя широкий или средний охват аудитории при низкой степени персонализации. К таким кампаниям по управлению восприятием относятся самые распространенные сейчас инструменты в виде «бот-ферм», генерирующих необходимую риторику в контексте конкретных событий [Bailurkar, Raul 2021]. В значительной степени целью таких кампаний должно стать смещение эмоционального режима всего общества в сторону большей допустимости ненависти к отдельным явлениям или группам. Так как степень персонализации в таких кампаниях низкая, ответом становится попытка платформ распознать «неаутентичное поведение» в сети и пресечь его на уровне модерации. К «неаутентичному» при этом относится результат активности не самих пользователей платформ, а ботов и основанных на ИИ систем.

Второй, более усовершенствованный сценарий подразумевает меньший охват аудитории при большей степени персонализации. В таких ситуациях качество эмоционального контента увеличивается, так как учитывает личностные характеристики пользователей. В данном случае подключается психологический профайлинг (т.е. составление психологического портрета пользователей на основе цифровых следов) отдельных социальных групп, предрасположенных к определенному контенту и информации. Для этого используются инструменты таргетинга (индивидуального подбора отображаемой информации) платформы. Одним из наиболее показательных примеров политического использования таргетинга стал скандал с Cambridge Analytica, проявившийся в ходе президентской кампании в США в 2016 г. Аналогичную ситуацию можно спрогнозировать и в сфере распространения ненависти со стороны других заинтересованных акторов. Опасения по поводу манипуляций в рамках политической онлайн-рекламы заставили множество стран, а также надгосударственных акторов начать разработку соответствующих законодательных ограничений для платформ в дополнение к уже существующим практикам контроля политической рекламы [Furnémont, Kevin 2020]. Среди них существуют страны, принявшие нормы, призванные повысить прозрачность рекламы для пользователей (Франция), а также нацеленные на применение правил по защите персональных данных (Великобритания) [Helberger, Dobber, de Vreese 2021]. Одной из наиболее масштабных инициатив в этой сфере можно считать проект Европейской комиссии, нацеленный на прозрачность политического таргетинга в преддверии выборов в Европарламент в 2024 г.⁴

Третий сценарий является наиболее сложным в реализации на текущий момент, но возможным в будущем. Речь идет о крупномасштабных попытках влия-

⁴ *European Commission*. European Democracy: Commission sets out new laws on political advertising, electoral rights and party funding. Press release, 25.11.2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6118 (дата обращения: 08.06.2022).

ния на эмоциональный режим всего общества посредством децентрализованного, но синхронного воздействия на большую часть населения через персонализированные и эмоционально оптимизированные сообщения в социальных сетях. По сути, такой подход подразумевает совмещение охвата первого сценария с практической реализацией второго. Подобная ситуация требует подключения мощных ресурсов, возможных к реализации лишь при применении искусственного интеллекта. Ключевым аспектом должна стать автоматизация создания, персонализации и распространения контента, способного изменить мнение людей, создав эмоциональный контекст, толкающий их к вражде [Kolotaev 2021]. Несмотря на все сложности, технический прогресс в сфере ИИ демонстрирует, что такой сценарий весьма вероятен в среднесрочной перспективе. При таком сценарии ответные действия должны осуществляться на двух уровнях. На техническом уровне требуется разработка контрмер, способных вычислять неаутентичный контент и, основываясь на сетевом анализе, вычислять структуры распространения этого контента.

Вместе с тем более сложным в реализации, но более результативным на практике может стать выработка цифровой грамотности населения. К этому процессу относятся образовательные инициативы⁵, а также действия гражданского общества и бизнеса по популяризации механизмов модерации и выдачи контента⁶. Такие инициативы имеют длительный процесс воплощения, что обуславливает необходимость дополнять их другими инициативами.

Таким образом, дальнейшее развитие цифровых платформ и технических инструментов, формирующихся уже сейчас, может спровоцировать более совершенные процессы в сфере распространения риторики ненависти. В то же время существуют различные механизмы противодействия, применимые к каждому из сценариев. Важно отметить, что их применение должно быть не ситуационным, а непрерывным, так как цифровая среда требует постоянного реагирования на угрозы. Кроме того, методы противодействия, рассчитанные на долгосрочную перспективу, требуют более длительной реализации, что также обуславливает необходимость их постоянного применения и совершенствования.

Выводы

Цифровая риторика ненависти демонстрирует, как новые технологии могут обострять проблемы, свойственные общественной и политической коммуникации. В политическом контексте особенно важно воздействие риторики ненависти на трансформацию эмоциональных режимов. Теоретическая концептуализация, представленная в работе, демонстрирует, в силу каких общественных механизмов язык вражды способен мотивировать широкую общественность или

⁵ Council of Europe. No Hate Speech Online Training Course. URL: <https://www.coe.int/en/web/youth/no-hate-speech-online-training-course> (дата обращения: 07.06.2022).

⁶ Code of Conduct on Countering Illegal Hate Speech Online. 2016. URL: https://ec.europa.eu/info/files/code-conduct-countering-illegal-hate-speech-online_en (дата обращения: 07.06.2022).

отдельные группы к вражде или насилию. С практической точки зрения, ключевым звеном в процессе разжигания ненависти становятся платформы, играющие роль и инструмента, и посредника. Эта взаимосвязанность требует сбалансированного ответа, выраженного в применении технических и социальных механизмов, смягчающих последствия цифровизации политических угроз. При этом важно учитывать не только реализующийся сценарий применения языка вражды, но и долгосрочные факторы, способствующие адаптации механизмов под эволюцию самой угрозы.

Поступила в редакцию / Received: 19.04.2022

Доработана после рецензирования / Revised: 11.06.2022

Принята к публикации / Accepted: 15.06.2022

Библиографический список

- Золян С.Т.* и др. Фейки: коммуникация, смыслы, ответственность / под ред. Г.Л. Тульчинского. СПб.: Алетейя, 2021.
- Мартьянов Д.С.* и др. Управляемость и дискурс виртуальных сообществ в условиях политики постправды / под ред. Д.С. Мартьянова. СПб.: ЭлекСис, 2019
- Попова О.В.* и др. «Политика постправды» и популизм / под ред. О.В. Поповой. — СПб.: Скифия-принт, 2018.
- Субочев В.В.* Политико-правовое манипулирование как основа управления обществом в эпоху постправды // Вестник Томского государственного университета. Право. 2019. № 34. С. 29–43.
- Хлопотунов Я.Ю.* Риторика ненависти в американском политическом дискурсе (функционально-лингвистический анализ) // Дискурс профессиональной коммуникации. 2020. Т. 2, № 2. С. 20–30.
- Ali K., Zain-ul-abdin K.* Post-Truth Propaganda: Heuristic Processing of Political Fake News on Facebook⁷ During the 2016 US Presidential Election // *Journal of Applied Communication Research*. 2021. Vol. 49, no. 1. P. 109–128.
- Bailurkar R., Raul N.* Detecting Bots to Distinguish Hate Speech on Social Media // 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021. P. 1–5.
- Bufacechi V.* Truth, Lies and Tweets: A Consensus Theory of Post-Truth // *Philosophy & Social Criticism*. 2021. Vol. 47, no. 3. P. 347–361.
- Castaño-Pulgarín S.A.* et al. Internet, Social Media and Online Hate Speech. Systematic Review // *Aggression and Violent Behavior*. 2021. Vol. 58. P. 101608.
- Chetty N., Alathur S.* Hate Speech Review in the Context of Online Social Networks // *Aggression and Violent Behavior*. 2018. Vol. 40. P. 108–118.
- Colwell T.M.* I.2 Emotives and Emotional Regimes // *Early Modern Emotions*. Routledge, 2016. P. 45–47.
- Cosentino G.* Social Media and the Post-Truth World Order. London: Palgrave Pivot, 2020.
- d'Ancona M.* Post-truth: The New War on Truth and How to Fight Back. Random House, 2017.
- Flaxman S., Goel S., Rao J.M.* Filter Bubbles, Echo Chambers, and Online News Consumption // *Public opinion quarterly*. 2016. Vol. 80, no. S1. P. 298–320.

⁷ 21 марта 2022 г. Тверской суд города Москвы признал Meta (продукты Facebook и Instagram) экстремистской организацией.

- Foucault M.* Subjectivity and Truth: Lectures at the Collège de France, 1980–1981: Michel Foucault, Lectures at the Collège de France / M. Foucault; eds. F. Ewald, A. Fontana, F. Gros. London: Palgrave Macmillan, 2017.
- Furnémont J.-F., & Kevin D.* Regulation of Political Advertising: A Comparative Study with Reflections on the Situation in South-East Europe. Council of Europe, 2020.
- Gagliardone I. et al.* Countering Online Hate Speech. UNESCO Publishing, 2015.
- Helberger N., Dobber T., de Vreese C.* Towards Unfair Political Practices Law: Learning lessons from the regulation of unfair commercial practices for online political advertising // Journal of Intellectual Property, Information Technology and E-Commerce Law. 2021. Vol. 12.
- Kalpokas I.* A Political Theory of Post-Truth. London: Palgrave Macmillan, 2019.
- Kapantai E. et al.* A Systematic Literature Review on Disinformation: Toward a unified taxonomical framework // New media & society. 2021. Vol. 23, no. 5. P. 1301–1326.
- Kolotaev Y.* Sentiment Analysis: Challenges to Psychological Security and Political Stability // ECIAIR 2021 3rd European Conference on the Impact of Artificial Intelligence and Robotics. Academic Conferences and publishing limited, 2021. P. 8–2-89.
- Lyman P.* The Domestication of Anger: The Use and Abuse of Anger in Politics // European Journal of Social Theory. 2004. Vol. 7, no. 2. P. 133–147.
- Matamoros-Fernández A., Farkas J.* Racism, Hate Speech, and Social Media: A Systematic Review and Critique // Television & New Media. 2021. Vol. 22, no. 2. P. 205–224.
- McGrath L.S.* Historiography, Affect, and the Neurosciences // History of Psychology. 2017. Vol. 20, no. 2.
- Morgan S.* Fake News, Disinformation, Manipulation and Online Tactics to Undermine Democracy // Journal of Cyber Policy. 2018. Vol. 3, no. 1. P. 39–43.
- Olteanu A. et al.* The Effect of Extremist Violence on Hateful Speech Online // Proceedings of the International AAAI Conference on Web and Social Media. 2018. Vol. 12, no. 1.
- Pinker S.* The Better Angels of Our Nature: The decline of violence in history and its causes. London: Penguin Books, 2011.
- Pitruzzella G., Pollicino O.* Disinformation and Hate Speech: A European Constitutional Perspective. Milano: Bocconi University Press, 2020.
- Prozorov S.* Why is There Truth? Foucault in the Age of Post-Truth Politics // Constellations. 2019. Vol. 26, no. 1. P. 18–30.
- Reddy W.M.* The Navigation of Feeling: A Framework for the History of Emotions. Cambridge: Cambridge University Press, 2001.
- Sellars A.* Defining Hate Speech // Berkman Klein Center Research Publication. 2016. No. 20. P. 16–48.
- Siddiquee M.A.* The Portrayal of the Rohingya Genocide and Refugee Crisis in the Age of Post-Truth Politics // Asian Journal of Comparative Politics. 2020. Vol. 5, no. 2. P. 89–103.
- Sullivan E.* The History of the Emotions: Past, Present, Future // Cultural History. 2013. Vol. 2, no. 1. P. 93–102.
- Süselbeck J.* Sprache und emotionales Gedächtnis. Zur Konstruktion von Gefühlen und Erinnerungen in der Literatur und den Medien // Emotionen. JB Metzler, Stuttgart. 2019. P. 282–295.
- Wahl-Jorgensen K.* Media Coverage of Shifting Emotional Regimes: Donald Trump's Angry Populism // Media, Culture & Society. 2018. Vol. 40, no. 5. P. 766–778.
- Wilber K.* Trump and a Post-Truth World. Shambhala Publications, 2017.

References

- Ali, K., & Zain-ul-abdin, K. (2021). Post-truth propaganda: Heuristic processing of political fake news on Facebook⁸ during the 2016 US presidential election. *Journal of Applied Communication Research*, 49(1), 109–128.
- Bailurkar, R., & Raul, N. (2021). Detecting bots to distinguish hate speech on social media. In *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 1–5.
- Bufacchi, V. (2021). Truth, lies and tweets: A consensus theory of post-truth. *Philosophy & Social Criticism*, 47(3), 347–361.
- Castaño-Pulgarín, S.A., Suárez-Betancur, N., Vega, L.M. T., & López, H.M. H. (2021). Internet, social media and online hate speech: Systematic review. *Aggression and Violent Behavior*, 58, 101608.
- Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and violent behavior*, 40, 108–118.
- Colwell, T.M. (2016). I. 2 Emotives and emotional regimes. In *Early Modern Emotions*. Routledge. 45–47.
- Cosentino, G. (2020). *Social Media and the Post-Truth World Order*. London; Cham: Palgrave Pivot.
- d’Ancona, M. (2017). *Post-truth: The New War on Truth and How to Fight Back*. Random House.
- Flaxman, S., Goel, S., & Rao, J.M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1), 298–320.
- Foucault, M. (2017). *Subjectivity and Truth: Lectures at the Collège de France, 1980–1981: Michel Foucault, Lectures at the Collège de France*. Palgrave Macmillan.
- Furnémont, J.-F., & Kevin, D. (2020). *Regulation of Political Advertising: A comparative study with reflections on the situation in South-East Europe*. Council of Europe.
- Gagliardone, I., et al. (2015). *Countering Online Hate Speech*. UNESCO Publishing.
- Helberger, N., Dobber, T., & de Vreese, C. (2021). Towards unfair political practices law: Learning lessons from the regulation of unfair commercial practices for online political advertising. *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 12
- Kalpokas, I. (2019). *A Political Theory of Post-Truth*. London: Palgrave Macmillan.
- Kapantai, E., et al. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New media & society*, 23(5), 1301–1326.
- Khlopotunov, Y.Y. (2020). Hate speech in American political discourse: Functional-linguistic analysis. *Professional Discourse & Communication*, 2(2), 20–30 (In Russian).
- Kolotaev, Y. (2021). Sentiment analysis: Challenges to psychological security and political stability. In *ECAIR 2021 3rd European Conference on the Impact of Artificial Intelligence and Robotics*. (pp. 82–89). Academic Conferences and publishing limited.
- Lyman, P. (2004). The domestication of anger: The use and abuse of anger in politics. *European Journal of Social Theory*, 7(2), 133–147.
- Martyanov, D.S., Bykov, I.A., Lukyanova, G.V., Martyanova, N., Rubtsova, M.V., & Podlesskaya, N.S. (2019). *Manageability and Discourse of Virtual Communities in the Context of Post-Truth Politics*. ElecSys. (In Russian).
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, 22(2), 205–224.
- McGrath, L.S. (2017). Historiography, affect, and the neurosciences. *History of Psychology*, 20(2).
- Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, 3(1), 39–43.

⁸ 21 марта 2022 г. Тверской суд города Москвы признал Meta (продукты Facebook и Instagram) экстремистской организацией.

- Olteanu, A., et al. (2018). The effect of extremist violence on hateful speech online. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1).
- Pinker, S. (2011). *The Better Angels of Our Nature: The decline of violence in history and its causes*. London: Penguin Books.
- Pitruzzella, G., & Pollicino, O. (2020). *Disinformation and Hate Speech: A European Constitutional Perspective*. Milano: Bocconi University Press.
- Popova, O.V., et al. (2018). *“Post-Truth Politics” and Populism*. St. Petersburg: Scythia-print. (In Russian).
- Prozorov, S. (2019). Why is there truth? Foucault in the age of post-truth politics. *Constellations: An International Journal of Critical and Democratic Theory*, 26(1). 18–30.
- Reddy, W.M. (2001). *The Navigation of Feeling: A Framework for the History of Emotions*. Cambridge University Press.
- Sellars, A. (2016). *Defining Hate Speech*. Berkman Klein Center Research Publication, (2016–20), 16–48.
- Siddiquee, M.A. (2020). The portrayal of the Rohingya genocide and refugee crisis in the age of post-truth politics. *Asian Journal of Comparative Politics*, 5(2), 89–103.
- Subochev V. (2019). Political-legal manipulation as the basis for the governance of society in the era of post-truth. *Tomsk State University Journal. Law*, (34), 29–43. (In Russian).
- Sullivan, E. (2013). The history of the emotions: Past, present, future. *Cultural History*, 2(1), 93–102.
- Süselbeck, J. (2019). Sprache und emotionales Gedächtnis. Zur Konstruktion von Gefühlen und Erinnerungen in der Literatur und den Medien. In J.B. Metzler, *Emotionen* (pp. 282–295). Stuttgart. (In German).
- Wahl-Jorgensen, K. (2018). Media coverage of shifting emotional regimes: Donald Trump’s angry populism. *Media, Culture & Society*, 40(5), 766–778.
- Wilber, K. (2017). *Trump and a Post-Truth World*. Shambhala Publications.
- Zolyan, S.T., Probst, N.A., Śladkiewicz, Ż., & Tulchinsky, G.L. (2021). *Fake: Communication, Meanings and Responsibility*. Independent Alliance. St. Petersburg: Aleteyya. (In Russian).

Сведения об авторе:

Колотаев Юрий Юрьевич — аспирант факультета международных отношений Санкт-Петербургского государственного университета (e-mail: yury.kolotaev@mail.ru) (ORCID: 0000-0001-8372-1193)

About the author:

Yury Y. Kolotaev — postgraduate of the School of International Relations of Saint Petersburg State University, Saint Petersburg, Russian Federation (e-mail: yury.kolotaev@mail.ru) (ORCID: 0000-0001-8372-1193)