

**А.М. Елеусизова**

Евразийский национальный университет имени Л.Н. Гумилева, Казахстан, г. Астана  
e-mail: aygelova\_altynay@mail.ru

## РЕКОМЕНДАЦИИ ВРЕДНОСНОГО КОНТЕНТА В YOUTUBE: МЕТА-АНАЛИЗ

Алгоритмы рекомендательной системы YouTube становятся объектом споров и дискуссий, отмечены инциденты с предложениями пользователям вредоносного, чаще всего экстремистского контента. Цель настоящего исследования – проверка гипотезы, что рекомендации YouTube могут содержать вредоносный контент, а также способствовать его распространению.

Научная и практическая значимость данной работы заключается в том, что сеть YouTube имеет огромную популярность, и знание об алгоритмах работы её рекомендательной системы поможет повысить информационную грамотность пользователей, и, как следствие, предостеречь от опасности, которую может вызвать воздействие на сознание вредоносной информации. Основная задача данного исследования заключается в популяризации идеи, что вся информация, размещённая в сети Internet, требует тщательной проверки на правдивость, а также, ввиду повышения информационной грамотности пользователей, настоящее исследование в глобальном плане способствует предотвращению терактов, актов членовредительства, самоубийства, педофилии и т. п.

Материалами для проведения исследования послужили работы, опубликованные за последние 5 лет, содержащие описание хотя бы одного из типов вредоносного контента. Поиск интересующего материала проводился на основе авторитетных наукометрических баз данных Google Scholar, Scopus, Web of Science и PubMed методом экстрагирования подходящих материалов. В результате, в соответствии с критериями приемлемости и исключения, было отобрано 22 исследования, после чего был проведён мета-анализ, по результатам которого установлено, что в 13 из них рекомендации YouTube содержали и способствовали распространению вредоносного контента, в 7 исследованиях учёные представили неоднозначные результаты, и только в 2 исследованиях авторами установлено, что рекомендации не содержали вредоносный контент и не способствовали его распространению.

Таким образом, согласно результатам данного исследования установлено, что в рекомендациях YouTube может содержаться и распространяться запрещённый вредоносный контент, в связи с чем авторы настоятельно рекомендуют внести коррективы в алгоритмы работы рекомендаций с целью ограждения пользователей от запрещённой информации, включая пропаганду экстремизма и насилия.

**Ключевые слова:** YouTube, рекомендации, рекомендательная система, вредоносный контент, псевдонаучный контент, радикальный контент, экстремизм, педофилия, мета-анализ.

А.М. Yeleussizova

L.N. Gumilyov Eurasian National University, Kazakhstan, Astana  
e-mail: aygelova\_altynay@mail.ru

### YouTube radical content recommendations: a meta-analysis

YouTube's recommendation system algorithms have been the subject of controversy and debate, as there have been incidents of the recommendation system offering hazardous, most often extremist content to users. The purpose of this study is to test the hypothesis that YouTube recommendations can contain and promote radical content.

The scientific and practical significance of this work is that YouTube has a huge popularity, and knowledge about the algorithms of its recommendation system would help to improve information literacy of users and thus prevent the potential danger that can be caused by exposure to malicious information. This study seeks to popularize the idea that all information on the Internet requires a thorough verification. Improving information literacy of users, this research contributes globally to the prevention of terrorist attacks, acts of self-mutilation, suicide, pedophilia, etc.

The sources for the present investigation with description at least one type of radical content were published in the last 5 years. The material of interest was searched through authoritative scientific and

metric databases, namely: Google Scholar, Scopus, Web of Science, and PubMed by extracting suitable records. A total of 22 studies were pooled according to the eligibility and exclusion criteria, and a meta-analysis was performed which found that in 13 of them YouTube recommendations contained and promoted hazardous content, in 7 studies the scientists presented ambiguous results, and only in 2 studies the authors found that the recommendations did not contain or promote any radical content.

Thus, according to the results of this research, it was found that YouTube recommendations can contain and disseminate prohibited hazardous content, and therefore the authors strongly recommend making adjustments to the algorithms of the recommendations in order to protect users from prohibited information, including the promotion of extremism and violence.

**Key words:** YouTube, recommendations, recommendation system, hazardous content, pseudoscientific content, radical content, extremism, pedophilia, meta-analysis.

А.М. Елеусизова

А.Н. Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Астана қ.  
e-mail: aygelova\_altynay@mail.ru

### Youtube – тегі зиянды контент бойынша ұсыныстар: мета – талдау

YouTube-тің ұсыныс жүйесінің алгоритмдері даулар мен пікірталастардың объектісіне айналады, өйткені зиянды, көбінесе экстремистік мазмұнды пайдаланушыларға ұсыныс жүйесінің нұсқаулығымен болған оқиғалар атап өтіледі. Бұл зерттеудің мақсаты – YouTube ұсыныстарында зиянды мазмұн болуы мүмкін, сонымен қатар оның таралуына ықпал етеді деген гипотезаны тексеру. Бұл жұмыстың ғылыми және практикалық маңыздылығы – YouTube желісі өте танымал және оның ұсыныс жүйесінің алгоритмдері туралы білу пайдаланушылардың ақпараттық сауаттылығын арттыруға көмектеседі, сонымен қатар зиянды ақпараттың санаға әсерін тудыруы мүмкін.

Бұл зерттеудің негізгі міндеті – Интернет желісінде орналастырылған барлық ақпараттың шынайылығы мұқият тексеруді қажет етеді, сонымен қатар пайдаланушылардың ақпараттық сауаттылығының артуына байланысты бұл зерттеу жаһандық деңгейде террористік актілердің, өзіне зиян келтіру, суицид, педофилия және т. б. теріс әрекеттердің алдын алуға ықпал етеді деген идеяны насихаттау. Зерттеу материалдары соңғы 5 жылда жарияланған және зиянды мазмұнның кем дегенде бір түрінің сипаттамасын қамтитын жұмыстар болды. Қызығушылық тудыратын материалды іздеу Google Scholar, Scopus, Web of Science және PubMed-тің беделді ғылыми-метрикалық мәліметтер базасы негізінде тиісті материалдарды алу әдісімен жүргізілді.

Нәтижесінде, жарамдылық және алып тастау критерийлеріне сәйкес 22 зерттеу таңдалды, содан кейін мета-талдау жүргізілді, оның нәтижелері бойынша YouTube ұсынымдарының 13- зиянды мазмұнды қамтитыны және таратуға ықпал еткені анықталды, 7 зерттеуде ғалымдар аралас нәтижелер ұсынды және тек 2 зерттеуде авторлар «ұсыныстарда зиянды мазмұн жоқ және оның таралуына ықпал еткен жоқ» деген қорытындыға келді.

Осылайша, осы зерттеудің нәтижелеріне сәйкес YouTube ұсынымдарында тыйым салынған зиянды мазмұн болуы және таралуы мүмкін екендігі анықталды, осыған байланысты авторлар пайдаланушыларды экстремизм мен зорлық-зомбылықты насихаттауды қоса алғанда, тыйым салынған ақпараттан қорғау мақсатында ұсынымдардың жұмыс алгоритмдеріне түзетулер енгізуді ұсынады.

**Түйін сөздер:** YouTube, ұсыныстар, кеңес беру жүйесі, зиянды мазмұн, псевдоғылыми мазмұн, радикалды мазмұн, экстремизм, педофилия, мета-анализ.

### Введение

Социальные сети, включающие миллиарды пользователей в глобальном информационном пространстве, стали неотъемлемой частью повседневной жизни, удовлетворяя потребности в общении и познании нового, при этом чаще всего они используются в развлекательных целях.

В работах казахстанских и зарубежных учёных неоднократно была доказана способность социальных медиа изменять мировоззре-

ние, жизненные ценности и влиять на интересы пользователей, формировать их жизненные ориентиры (Есенбекова, 2022: 53; Green, 2019; Roose, 2019; Allcott et al., 2020: 629-676; Schaub & Morisi, 2020: 752-773; Williams et al., 2020: 93-117; Muller & Schwarz, 2021: 2131-2167), в связи с чем огромное беспокойство вызывает тот факт, что социальные сети могут служить источником fake news, пропаганды насилия, педофилии, содержать призывы к экстремизму с неблагоприятными последствиями для общества.

Согласно понятию, изложенному в «Оксфордском словаре», экстремизм (лат. «extremus» – «крайний», «чрезвычайный») – это политические, религиозные и другие действия, являющиеся неприемлемыми для большинства людей. Отметим, что социальные сети являются эффективным инструментом распространения идей экстремизма во всём мире, в том числе и в Республике Казахстан.

Одним из действенных способов противодействия пропаганде терроризма и экстремизма в Internet-пространстве является не только фильтрация запросов, блокирование Internet-сайтов и своевременное удаление противозаконной информации, но и повышение религиозной грамотности современной молодёжи посредством пропаганды Internet-ресурсов Духовного управления мусульман Казахстана и духовно-просветительских сайтов «E-islam.kz» и «Islam.kz» (Абдрашев, 2016: 59).

Высокую опасность влечёт распространение идей ваххабизма – религиозно-политической исламской секты. Основная опасность заключается в том, что те ценности, которые они пропагандируют, привлекают современную молодёжь. Если в прошлом идеи исламского радикализма получали своё распространение методом изготовления в тайных типографиях и распространения листовок с призывами свержения конституционного строя нашей Республики, то в период активного использования сети Internet задача ваххабитов намного упростилась. При этом вероисповедание или владение определённым языком не являются решающим фактором. Так, в своём исследовании Абдрашев (2016) сообщает, что русскоязычное население, наравне с казахоязычным, вовлекается в ваххабитские сообщества. При этом автор отмечает специфическую особенность распространения религиозных идей, заключающуюся в использовании их замаскированных форм, к которым относится, помимо прочих, насильственное свержение конституционного строя нашей страны. Вербовка осуществляется преимущественно посредством распространения видеороликов экстремистского содержания по сети YouTube, после просмотра которых пользователи вовлекаются в дискуссии, оставляя комментарии к просмотренным материалам. Тот факт, что распространение идей экстремизма носит явно открытый характер, подчеркивает транснациональный характер организованных преступных формирований (Абдрашев, 2016: 58-62).

Так, в конце 2022 года Генпрокуратура Казахстана заблокировала веб-ссылку видеоконтента Youtube (около 3 000 подписчиков и 5 000 просмотров) лжепроповедника, причастного к международной террористической организации.

При этом уточним: Министерство информации и общественного развития Казахстана сообщает о том, что администрация YouTube и Twitter не удаляют запрещённый контент по их просьбе: представители интернет-компаний в целом с пониманием воспринимают предложения о сотрудничестве в вопросах удаления противоправного контента, нарушающего требования международного права и законодательства Республики Казахстан, и в рамках практического взаимодействия удаляют материалы, содержащие пропаганду экстремизма, терроризма, порнографии, суицида, онлайн-продажи наркотиков и подобные, однако YouTube и Twitter, к сожалению, всё ещё не готовы к сотрудничеству.

Актуальность темы настоящего исследования выражена, в первую очередь тем, что социальные сети, включая YouTube, недостаточно урегулированы с точки зрения правовых норм Казахстана. К примеру, активный рост числа пользователей YouTube, в том числе несовершеннолетних, общение посредством комментариев, доступность видеоматериалов различного характера, привлекают пристальное внимание общественности к проблемам регулирования, выработке правил справедливого, легитимного функционирования данной социальной сети со стороны государства.

Роль YouTube в облегчении доступа к вредоносному контенту посредством рекомендательной системы – вопрос дискуссионный, требующий детального изучения. Цель настоящего исследования – установить, способствует ли рекомендательная система YouTube распространению вредоносного контента. Нас интересует сценарий, при котором алгоритмы YouTube рекомендуют вредоносный контент при просмотре безобидного видео, канала, или при просмотре видео, канала, содержащего вредоносный контент. Под «вредоносным контентом» мы подразумеваем контент, который нарушает принципы сообщества, установленные YouTube (материалы сексуального характера: видео с демонстрацией обнажённых тел, актами педофилии; самоубийство и членовредительство, а также контент, содержащий насильственные и опасные материалы: разжигание ненависти, экстремистский контент, контент агрессивных преступных организаций и дезинформация).

## Обзор литературы

Исчерпывающее количество исследований указывает на то, что вредоносный контент становится для пользователей всё более доступным. В ряде исследований рассматривается доступность соответствующего контента на YouTube и роли рекомендательной системы данной сети в облегчении распространения экстремистского контента (Kaakinen et al., 2018: 90-97; Schmitt et al., 2018: 708-808; Spinelli & Crovella, 2020: 244-251; Alfano et al., 2021: 835-858; Nienierza et al., 2021: 1229-1246; Chen et al., 2022). Результаты исследований отвечают на вопрос, ищут ли пользователи вредоносный контент самостоятельно, или он предлагается пользователям посредством рекомендательной системы. Верно второе утверждение.

В исследовании Kaakinen et al. (2018) сообщается, что вероятность того, что пользователи в возрасте от 15 до 30 лет столкнутся с ненавистническим контентом, связанным с полом, этнической принадлежностью, политическими взглядами, терроризмом и религией в социальных сетях, за последние 5 лет утроилась (Kaakinen et al., 2018: 90-97). Исследования той же возрастной группы, проведённые Nienierza et al. (2021) установили, что 40% респондентов сообщили о том, что наблюдали экстремистский контент в социальных сетях (Nienierza et al., 2021: 1229-1246).

Можно ошибочно предположить, что пользователи намеренно ищут контент такого плана, просматривая видеоматериалы YouTube. Так, Munger & Phillips (2022) сообщают, что пользователи ищут интересующий их контент, а создатели контента лишь удовлетворяют их любопытство (Munger & Phillips, 2022: 186-219). Однако данный подход игнорирует тот факт, что алгоритм YouTube по умолчанию автоматически воспроизводит дополнительные видео, которые его рекомендательная система считает интересными для пользователя. Даже в том случае, когда функция автовоспроизведения отключена, пользователям предлагаются видео на боковой панели.

Наряду с рекомендательной системой, сеть YouTube в основном состоит из отдельных видеоматериалов и каналов. Канал – это уникальное пространство определённого пользователя на YouTube, куда он может загружать видеоролики, а другие пользователи могут с ними ознакомиться в случае, если видеоматериал отмечен,

как «общедоступный». YouTube предоставляет рекомендации по каналам, а также по отдельным видеороликам. Алгоритмы рекомендаций YouTube анализируют активность пользователя во взаимосвязи с производителями видеоматериалов, чтобы предлагать или автоматически воспроизводить видео пользователям. Следовательно, в отличие от неуправляемого выбора пользователей, рекомендательная система отвечает за 70% просмотров видео на YouTube. Таким образом, влияние рекомендательной системы YouTube на аудиторию вызвало вопросы о её роли в распространении вредоносного, запрещённого контента.

Так, Courtois & Timmermans (2018) сообщают, что рекомендательная система YouTube может направлять внимание пользователей на видеоконтент, про который в других случаях они бы даже не узнали (Courtois & Timmermans 2018: 1-16). К примеру, недавнее исследование Alfano et al. (2021) показало, что пользователи могут получить доступ к запрещённому контенту через рекомендательную систему из видеороликов о фитнесе, огнестрельном оружии, обучении, и даже о строительстве домов (Alfano et al., 2021: 835-858). Другие исследования, проведённые Röchert et al. (2020), выявили в сети YouTube так называемые «пузыри фильтров» (filter bubbles), которые основаны на идее релевантности, это алгоритмически контролируемое представление однородного контента пользователям (Röchert et al., 2020: 245-254). Такой контент обычно соответствует интересам, идеям и убеждениям пользователей.

Рекомендательная система YouTube из всплывающих окон с фильтрами не раз подвергалась критике. Так, имеется ряд исследований, в которых подчеркиваются потенциальные негативные последствия алгоритмов – акты насилия, спровоцированные соответствующими видеоматериалами. К примеру, в исследовании Green (2019) упоминается случай о 26-летнем мужчине из Сиэтла (США) по имени Buckley Wolfe, который убил своего брата ножом, так как под воздействием от просмотра определённых видео посчитал, что его брат – инопланетная рептилия, способная к трансформации (Green, 2019). Идея Buckley основывалась на теории заговора о том, что инопланетные рептилии живут среди людей, которую пропагандировал David Icke. Так, Green, проанализировав «лайки» Buckley на YouTube, отметил постепенное изменение интересов данного пользователя. Изначально он

просматривал видеоролики преимущественно о боевых искусствах и фитнесе, но в конечном итоге его интересы сместились в сторону Alt-lite (правое политическое движение, которое дистанцируется от этнического национализма, но выступает против политкорректности, феминизма и ислама), теорий заговора. Данный инцидент усилил опасения по поводу алгоритмов рекомендательной системы YouTube.

В исследовании Roose (2019) сообщается, что несколько пользователей столкнулись с экстремистским контентом на YouTube. К примеру, Caleb Cain, бывший экстремист из Западной Вирджинии (США), рассказал репортёру газеты «New York Times», как увлёкся идеей экстремизма, просматривая тысячи видеороликов, содержащих материалы по женоненавистничеству, расизму и теории заговора. Caleb отмечает, что перечисленный контент «зомбировал» его сознание таким образом, что он почувствовал себя частью группы людей, обладающих исключительными знаниями о социальных проблемах (Roose, 2019). Данный случай уникален, так как в настоящее время Caleb осознаёт, что заблуждался, более того, он активно рассказывает пользователям о негативном влиянии, которое оказала на его сознание рекомендательная система YouTube.

С ростом опасений по поводу реальных последствий социальных сетей увеличивается количество исследований, целью которых является выявление причинно-следственного воздействия социальных сетей на антидемократические взгляды и поведение пользователей. К примеру, исследование Schaub & Morisi (2020), проведённое в Германии и Италии, показало, что лица, имеющие широкий доступ к сети Internet на муниципальном уровне, с большей вероятностью проголосуют за популистские партии (Schaub & Morisi, 2020: 752-773).

Согласно исследованию Muller & Schwarz (2021), активные высказывания против беженцев в социальной сети Facebook предсказывают реальные преступления. При этом авторы отмечают, что количество преступлений против беженцев уменьшилось во время перебоев в работе Facebook и в целом Internet-сети в отдельных районах (Muller & Schwarz, 2021: 2131-2167).

Рандомизированный эксперимент Allcott et al. (2020), проведённый на выборке пользователей из США, показал, что деактивация Facebook за 4 недели до промежуточных выборов 2018 года минимизировала возможность пользовате-

лей узнавать фактические новости и распространять политическую поляризацию, что позволяет предположить причинно-следственный эффект влияния сети Facebook на политическую поляризацию (Allcott et al., 2020: 629-676).

На Западе ультраправые и некоторые популярные правые политики сыграли ключевую роль в смещении «окна Овертона» политических дискуссий в сети Internet в сторону радикализма, создав медиапространство, где язык ненависти стал нормой. Исследования показывают, что деятели ультраправого толка давно функционируют в англоязычной сети Internet, практически не встречая препятствий со стороны правоохранительных органов ввиду конституционной защиты свободы слова в США. Результатом стало создание экстремистских веб-анклавов, которые обеспечивают коллективную виртуальную идентичность для ранее разрозненных сеятелей ненависти к тем или иным группам людей. Эти очаги способствовали активизации радикальных группировок во многих странах, включая США, Великобританию, Германию, Нидерланды, Италию и Швецию (Williams et al., 2020: 93-117).

Так, социальные сети стали причастны к террористической атаке ультраправых экстремистов в Крайстчерче (Новая Зеландия) в марте 2019 года. Террорист был заядлым пользователем социальных сетей, включая Facebook, где он транслировал своё нападение в прямом эфире. В начале своего сообщения он заявил, что переходит от словоизлияний в сети Internet к активным действиям вне сети. Видеотрансляция на Facebook продолжалась 17 минут. Ролик был удалён с платформы в течение часа, но уже успел распространиться по сети, был повторно загружен более 2 миллионов раз на Facebook, YouTube, Instagram и Twitter, и оставался доступным ещё сутки после террористического акта. Комментарии наводнили восторженные отзывы и горячая поддержка атаки (Williams et al., 2020: 93-117).

Вышеперечисленные исследования указывают на то, что социальные сети могут оказывать негативное влияние на политические взгляды пользователей и, таким образом, вызывают дополнительную обеспокоенность. Следовательно, в обществе возникают вопросы об ответственности социальных сетей и роли, которую они играют в преступлениях на почве ненависти и разобщённости.

Согласно анализу социальных сетей, националистических групп, транслирующих идею

«Казахстан – для казахов» не обнаружено, так как важнейшими чертами казахстанского народа на протяжении многих лет остаются способность к сочувствию, сопереживанию, добродушие, развитое чувство юмора, недостаточная прагматичность, уступчивость, великодушие, веротерпимость, гостеприимство, высокий творческий потенциал, изобретательность, трудолюбие, приверженность традициям, чувство долга, вежливость, почитание и уважение старших [Абдрашев, 2016: 61].

Коллектив казахстанских учёных во главе с Есенбековой (2022) при исследовании влияния социальных медиа на активность молодёжи, установили склонность казахстанской молодёжи к участию в конкретной общественной деятельности посредством форм выражения субъективных мнений в социальных сетях. При этом авторы отмечают, что социальная активность молодёжи обязана служить интересам государства и общества (Есенбекова, 2022: 53).

Разработчики YouTube признают потенциальное неблагоприятное воздействие сети и предпринимают шаги для сокращения вредоносного контента путём внесения изменений в алгоритмы системы. Они обязуются отслеживать и удалять вредоносный контент, продвигать в рейтинге авторитетные голоса с целью распространения проверенной информации, вознаграждать доверенных авторов и сокращать распространение контента, который противоречит политике. Так, в период с октября по декабрь 2020 года представители YouTube сообщили, что они удалили 3,8 миллиона видеороликов, нарушающих правила безопасности детей, 1,4 миллиона спама, вводящего пользователей в заблуждение, включая мошеннические видеоматериалы, 259 тысяч вредоносных или оскорбительных видео, 73 тысячи видеороликов, пропагандирующих насилие или насильственный экстремизм. Вышеуказанные меры указывают на то, что разработчики YouTube осведомлены о доступности вредоносного контента на своей платформе. Следует обратить внимание на необходимость проведения дальнейших исследований, которые смогут проверить эффективность принятых мер.

### Материалы и методы

Материалами для проведения исследования послужили работы, опубликованные за последние 5 лет, и содержащие описание хотя бы одного из типов вредоносного контента. Поиск

интересующего материала проводился на сайтах наукометрических баз данных Google Scholar, Scopus, Web of Science и PubMed методом экстрагирования подходящих материалов. В результате, в соответствии с критериями приемлемости и исключения, было отобрано 22 исследования.

В первую очередь выборка исследований производилась по названию, затем – по содержанию аннотации, а затем – путём прочтения полного текста научной работы. На каждом этапе процесса выборки исследования оценивались по критериям приемлемости, согласно которому исследование должно:

1. быть посвящено работе рекомендательной системы YouTube, либо анализировать алгоритмы YouTube наравне с другими объектами исследования. Исключались исследования, в которых изучалась рекомендательная система других социальных сетей (Instagram, Facebook, ВКонтакте, Twitter и т.п.);

2. быть сосредоточено на определённом канале, либо на видеорекомендациях. Также в настоящее исследование включались работы, рассматривающие вопросы создания классификаторов проблемного контента в контексте рекомендательной системы YouTube;

3. отвечать на вопрос, способствует ли рекомендательная система YouTube распространению вредоносного контента и облегчению его поиска;

4. быть сосредоточено на вредоносном содержании, включая экстремизм, разжигание ненависти, радикализацию, теории заговора, fake news;

5. быть опубликовано в рецензируемом журнале с доступом к полному тексту работы.

В случае, если научная работа не соответствовала критериям приемлемости, она исключалась из настоящего исследования. Так, были исключены исследования, посвящённые комментариям пользователей по теме экстремизма в отличных от YouTube социальных сетях, а также исследования, направленные исключительно на разработку системы обнаружения вредоносного контента. Дополнительно из первоначального корпуса статей были удалены дублирующиеся исследования (чаще всего, ввиду их размещения на нескольких платформах одновременно). Далее полученные данные были подвергнуты мета-анализу, который позволил выявить степень причастности рекомендательной системы YouTube к распространению вредоносного контента.

## Результаты и обсуждение

Согласно результатам проведённого мета-анализа, 13 из 22 исследований указали на то, что рекомендательная система YouTube способствует распространению, а также облегчает ли поиск вредоносного контента (Alfano et al., 2021: 835-858; AVAAZ Report, 2020; Chen et al., 2022; Fano et al., 2022: 779; Hussein et al., 2020: 1-27; Kaiser & Rauchfleisch, 2020: 2056305120969914; Kaiser et al., 2021: 1244-1262; Nickles et al., 2021: 920-923; Papadamou et al., 2020: 522-533; Papadamou et al.,

2021: 1-25; Röchert et al., 2020: 245-254; Schmitt et al., 2018: 780-808; Spinelli & Crovella, 2020: 244-251). По данным других двух исследований, рекомендательная система YouTube не способствовала распространению вредоносного контента (Hosseinmardi et al., 2020; Ledwich & Zaitsev, 2019). Остальные 7 из 22 исследований содержали неоднозначные результаты (Abul-Fottouh et al., 2020: 104175; Faddoul et al., 2020; Kaiser & Rauchfleisch, 2019: 1-22; Ledwich et al., 2022; Papadamou et al., 2022: 723-734; Ribeiro et al., 2020; Stöcker & Preuss, 2020: 359-375) (таблица 1).

Таблица 1 – Результаты проведённого мета-анализа

Исследование	Результаты исследования рекомендательной системы YouTube	Выводила ли рекомендательная система YouTube на проблемный контент?
1	2	3
Abul-Fottouh et al., 2020	YouTube рекомендовал больше нейтральных и про-вакцинных видео, чем анти-вакцинных. Видеоролики, агитирующие против вакцин, с меньшей вероятностью способствовали переходу к видеороликам, пропагандирующим применение прививок.	Неоднозначные результаты
Alfano et al., 2021	Продвижение теорий заговора при помощи таких поисковых запросов, как «natural foods», «gurus», и пр.	Да
AVAAZ Report, 2020	16% процентов топ-100 видео содержали дезинформацию об изменении климата и имели 21 миллион просмотров. Более 20% просмотров топ 100 роликов по поисковому запросу «global warming» содержали дезинформацию о климате.	Да
Chen et al., 2022	Рекомендательная система YouTube выводила просматривающих экстремистский контент на аналогичные видеоматериалы.	Да
Faddoul et al., 2020	Нет существенных доказательств наведения пользователя на правый контент. Отмечена высокая вероятность рекомендации конспирологических видео после просмотра конспирологических роликов.	Неоднозначные результаты
Fano et al., 2022	Рекомендательная система способствовала продвижению видеороликов с псевдонаучным медицинским контентом с домашней страницы при поиске видео на тему «синдактилия».	Да
Hosseinmardi et al., 2020	Отмечено растущее число эхо-камер ультраправых, однако не было обнаружено доказательств причастности к этому алгоритма рекомендаций.	Нет
Hussein et al., 2020	История просмотров пользователя влияла на количество дезинформации в рекомендациях YouTube. Эффект пузыря фильтров был выявлен по всем конспирологическим темам.	Да
Kaiser & Rauchfleisch, 2019	Видео интимного характера с участием несовершеннолетних были доступны через систему рекомендаций. Однако большинство из них находились на расстоянии десяти переходов. Видео были случайно обнаружены при анализе бразильских политических видео.	Неоднозначные результаты
Kaiser & Rauchfleisch, 2020	Выявлены ультраправые сообщества на YouTube – свидетельство эффекта пузыря фильтров.	Да
Kaiser et al., 2021	Алгоритм видеорекомендаций YouTube способствовал продвижению дезинформации о вирусе Зика.	Да
Ledwich & Zaitsev, 2019	Рекомендательная система не способствовала переходу к радикальному или экстремистскому контенту.	Нет

1	2	3
Ledwich et al., 2022	YouTube генерирует умеренные пузыри фильтров для большинства пользователей при доступе к экстремистскому контенту; ключевую роль играет персонализация пользователя.	Неоднозначные результаты
Nickles et al., 2021	При поиске видео на тему «лечение угревой сыпи» рекомендательная система в основном предлагает видеоролики с псевдонаучным медицинским контентом с домашней страницы.	Да
Papadamou et al., 2020	Дети, просматривающие видео для детей в возрасте от 1 до 5 лет, могут столкнуться с сексуальным контентом.	Да
Papadamou et al., 2021	При просмотре по 5 видеоматериалов из рекомендаций, пользователи увидят видео о сексуальном насилии.	Да
Papadamou et al., 2022	Пользователи чаще встречали псевдонаучные видео в результатах поиска, чем в рекомендациях или на главной странице. Кроме того, псевдонаучный контент в рекомендациях и на главной странице был представлен в незначительном количестве.	Неоднозначные результаты
Ribeiro et al., 2020	Рекомендательная система способствовала предложению контента от альтернативно правого политического крыла и тёмной сети интеллектуалов, но только через соответствующие тематические каналы.	Неоднозначные результаты
Röchert et al., 2020	Смешение право-популистских и политически нейтральных видеороликов на немецком языке.	Да
Schmitt et al., 2018	Видеоролики с контрпропагандой оказались сцеплены с экстремистским контентом: он содержался в результатах поисковых запросов, связанных с контрпропагандой.	Да
Spinelli & Crovella, 2020	Рекомендации YouTube перенаправляли пользователей с надёжного контента на видео, содержащие экстремистские и антинаучные точки зрения. Кроме того, пользователи, у которых личная информация была скрыта, получали больше рекомендаций к просмотру экстремистского и антинаучного контента.	Да
Stöcker & Preuss, 2020	Контекстуально некорректные рекомендации, которые были расценены как побочный эффект (в немалой степени вызванный автовоспроизведением), нежели цель рекомендательной системы.	Неоднозначные результаты

Исследования, которые рассмотрены в данной работе, описывают псевдонаучный контент, экстремистский контент, контент сексуального характера.

Рассмотрим группу исследований, по результатам которых установлено, что рекомендательная система может способствовать распространению вредоносного контента.

Исследователи Alfano et al. (2021) использовали веб-сканер для имитации потребления контента пользователями посредством рекомендательной системы. Цель исследования состояла в определении, предлагает ли рекомендательная система YouTube псевдонаучный контент при помощи таких поисковых запросов, как «natural foods», «gurus», и пр. По результатам исследования установлено, что пользователи могут столкнуться с контентом данного типа, просматривая вполне безобидные видео о боевых искусствах и натуральных продуктах (Alfano et al., 2021: 835-858).

В исследовании Hussein et al. (2020) использовали иную методологию. Авторы создали ботов и зарегистрировали на них учётные записи Google. У ботов, просматривавших видео с конспирологическим контентом, он стал предлагаться в рекомендациях YouTube (Hussein et al., 2020: 1-27). Полученные данные подтверждают эффект пузыря фильтров, указанный нами ранее, и описанный в работах Röchert et al. (2020).

Анализ социальных сетей, приведённый в отчёте AVAAZ (2020), показал, что 16% из 100 наиболее популярных видеороликов, найденных по поисковому запросу «глобальное потепление», содержали дезинформацию об изменении климата, являясь псевдонаучными. Согласно результатам анализа, просмотр вышеуказанных видеоматериалов в дальнейшем приводил к рекомендациям видео подобного характера (AVAAZ Report, 2020), что также указывает на эффект «пузыря фильтров».

С помощью сетевого анализа и тематического моделирования Kaiser et al. (2021) выявили кластеры дезинформации о вирусе Зика, продвигаемой с помощью алгоритма видеорекомендаций YouTube в Бразилии. Наиболее популярные видео на португальском и английском языках по данному топику содержали в основном точную и правильную информацию о вирусе Зика, но всё же изобиловали ошибочной и потенциально опасной дезинформацией, которая зачастую находилась всего в одном клике от изначального ролика и не была изолирована в конспирологическом пузыре фильтров, а находилась в дальнем конце алгоритма рекомендаций, то есть на каналах, которые YouTube рекомендовал, хоть и в меньшей степени. В то же время выявленные кластеры с наибольшим количеством просмотров относились к музыкальной тематике с ключевым словом Zika (в среднем более миллиона просмотров на один ролик). Видео, содержащие дезинформацию, значительно отставали по количеству просмотров (например, ролик против вакцинации от вируса Зика = 9 646 просмотров; видео, пропагандирующее мёд как лекарство от данного вируса = 120 491 просмотр). Однако это не отменяет факт наличия вредоносного явления (Kaiser et al., 2021: 1244-1262).

В исследовании Fano et al. (2022), рассматривающем дезинформацию в сфере медицинского контента, из найденных по ключевым словам рекомендуемых видео с домашней страницы по теме «синдактилия», только 14,8% видеороликов были созданы врачами, в 11,1% из них принимали участие сертифицированные хирурги-ортопеды. Остальные видеоролики могли содержать псевдонаучный контент (Fano et al., 2022: 779).

Подобный поиск по теме лечения угревой сыпи у подростков, проведённый Nickles et al. (2021), выявил, что 65,6% видео были основаны на точках зрения пациентов или блоггеров и не всегда содержали достоверную информацию; только 14,6% освещали мнение дерматологов; 8,3% видео отражали мнение представителей СМИ; 7,3% – врачей, не являющихся дерматологами; 2,1% – косметологов; и 2,1% видео – неустановленных личностей (Nickles et al., 2021: 920-923).

Исследователи Spinelli & Crovella (2020), смоделировали поведение пользователя, при этом использовался алгоритм для имитации просмотра пользователем контента на YouTube, а по окончании просмотра случайным образом выбиралось рекомендуемое видео для следующего

просмотра. Таким образом было установлено, что рекомендательная система может привести к экстремистскому и псевдонаучному контенту из надёжных источников информации (Spinelli & Crovella, 2020: 244-251).

Группа исследователей Schmitt et al. (2018) рассматривала возможность появления в рекомендациях экстремистского контента, и установила, что он тесно или напрямую связан с материалом, противоположным по теме (в данном случае, антиэкстремистскими материалами, предоставленными общественными организациями или правительством). К примеру, результаты показали, что пользователи могут столкнуться с экстремистским контентом в два клика через рекомендательную систему контента антиэкстремистских видеороликов (Schmitt et al., 2018: 780-808). Данный факт свидетельствует о том, что пользователи посредством рекомендательной системы легко могут столкнуться с вредоносным контентом, достаточно лишь совершить поиск диаметрально противоположного по смыслу запроса.

В исследовании, проведённом Chen et al. (2022), использовался подключаемый модуль браузера для отслеживания истории просмотров участников и алгоритмических рекомендаций YouTube. Результаты показали, что 9,2% участников просматривали видео экстремистского канала, 22,1% просматривали видео канала AIN («Альтернативная сеть влияния») – альтернативная медиасистема, стимулирующая онлайн-радикализацию. При просмотре соответствующих видеоматериалов, пользователям с большей вероятностью рекомендовались подобные видео. Таким образом, рекомендательная система YouTube может облегчить доступ к вредоносному контенту после того, как пользователь просматривает подобные видеоматериалы. Авторы отмечают, что 90% пользователей, просмотревших AIN и экстремистские видео, набрали высокие баллы в опросе по расовому недовольству, что свидетельствует о влиянии просмотренного контента на сознание пользователей (Chen et al., 2022).

Беспокойство вызывает исследование, проведённое Papadamou et al. (2020), в котором исследователи заключают, что посредством рекомендательной системы дети, просматривающие видео для детей в возрасте от 1 до 5 лет, могут столкнуться с сексуальным контентом (Papadamou et al., 2020: 522-533). Те же авторы провели другой подобный эксперимент для

установления влияния рекомендательной системы на просмотр контента, связанный с сексуальным насилием. Анализ случайного сёрфинга определил 18,8%-ную вероятность того, что пользователи, просмотрев по 5 видеоматериалов из рекомендаций, смогут увидеть хотя бы одно видео по соответствующей теме (Papadamou et al., 2021: 1-25).

Kaiser & Rauchfleisch (2020) в результате анализа рекомендаций каналов на YouTube, установили, что алгоритм рекомендации каналов YouTube способствует агломерации ультраправых сообществ в США ( $n = 13\,529$  каналов) и в Германии ( $n = 8000$  каналов). Отмечен эффект пузыря фильтров (Kaiser & Rauchfleisch, 2020: 2056305120969914).

В исследовании Röchert et al. (2020) проводят анализ однородности политического контента на основании просмотра 1663 соответствующих видеоматериалов. Авторы заключили, что шанс увидеть правопопулистский контент наравне с политически нейтральным контентом в рекомендательной системе достаточно высок (Röchert et al., 2020: 245-254). Полученные результаты могут послужить предварительными доказательствами роли рекомендаций YouTube в содействии к созданию информационных пространств для идеологических единомышленников.

Имеется малочисленное количество исследований, доказывающих непричастность рекомендательных алгоритмов YouTube к распространению вредоносной информации.

К примеру, для определения облегчения доступа к экстремистскому контенту Hosseinmardi et al. (2020) анализировали данные участников, хотя бы единожды посетивших YouTube, с целью выявления последовательных просмотров видео пользователями и рекомендации им подобных видеоматериалов. Согласно анализу, зарегистрировано растущее число эхо-камер ультраправых, однако доказательства того, что оно вызвано алгоритмом рекомендации, не установлено (Hosseinmardi et al., 2020).

В исследовании, проведённом Ledwich & Zaitsev (2019), был создан протокол для оценки количества показов рекомендаций пользователю, и сделан вывод о том, что рекомендательная система не облегчает доступ к экстремистскому контенту (Ledwich & Zaitsev, 2019).

Однако исследование, проведённое Ribeiro et al. (2020) установило связь рекомендательной системы с показами материалов Alt-lite и тёмной

сети интеллектуалов (IDW, от англ. Intellectual Dark Web Content – неформальная группа публичных деятелей, выступающих против политкорректности, но не обязательно поддерживающих экстремистскую идеологию). Однако, вышеуказанные материалы рекомендовались пользователю только в случае просмотра аналогичных видео, а при других сценариях они оказывались недоступными к просмотру (Ribeiro et al., 2020).

Таким образом, в данных двух аналогичных исследованиях были достигнуты различные результаты. Ledwich & Zaitsev (2019) предполагают, что за доступность вредоносного контента несут ответственность создатели контента, а не сеть YouTube. Они согласны с мнением Munger & Phillips (2019) о том, что YouTube – это структура, базирующаяся на спросе и предложении пользователей, которые ищут именно тот контент, в просмотре которого активно заинтересованы, а не пассивно переходят к вредоносному контенту от безобидных видеороликов.

Однако в недавнем исследовании Ledwich et al. (2022) получили неоднозначные результаты. Для анализа алгоритма рекомендаций YouTube при облегчении доступа к экстремистскому контенту авторы использовали ботов с предпочтениями контента и историей просмотров. Каждый бот сначала просматривал контент, соответствующий его запрограммированным предпочтениям, затем были показаны видеоролики из рекомендаций. Установлено, что YouTube генерирует умеренные пузыри фильтров для большинства пользователей. Однако, эффект пузыря фильтров для людей, посматривающих узкоспециализированный контент, значительно низкий. Политическим деятелям, за исключением пользователей из основных СМИ, рекомендуется меньше видео из категории контента основных СМИ, чем анонимным зрителям без персонализации. Также отмечено, что персонализация оказывает большее влияние на контент домашней страницы, чем на видео, рекомендуемые в ленте (Ledwich et al., 2022).

В исследовании Faddoul et al. (2020) провели лонгитюдный анализ с целью выяснить, с какой частотой алгоритмы YouTube рекомендуют псевдонаучный контент. В ходе анализа количество рекомендаций данного контента постепенно уменьшалось. Однако авторы доказали, что рекомендательная система может рекомендовать конспирологический контент, если пользователь просматривает аналогичные видео, под-

тверждая эффект пузыря фильтров. Однако нет существенных доказательств наведения пользователя на правый контент (Faddoul et al., 2020).

В аналогичном исследовании Abul-Fottouh et al. (2020) определяли частотность показов рекомендаций видеороликов «за» и «против» вакцинации от COVID-19. Результаты показали, что алгоритмы YouTube рекомендовали больше нейтральных и про-вакцинных видео, чем анти-вакцинных. При этом видеоролики, агитирующие против вакцин, с меньшей вероятностью способствовали переходу к видеороликам, пропагандирующим применение прививок (Abul-Fottouh et al., 2020: 104175).

Ранее упомянутые исследователи Papadamou et al. (2022) заключили, что пользователи чаще встречали псевдонаучные видео на странице результатов поиска YouTube, чем через рекомендательную систему или домашнюю страницу. Чаще всего пользователи сталкиваются с псевдонаучным контентом при поиске контента на темы против вакцинации от COVID-19, обязательного ношения масок и т. п., чем через рекомендательную систему. Кроме того, псевдонаучный контент в рекомендациях и на главной странице был представлен в незначительном количестве. Исследование также показало, что рекомендательная система с меньшей вероятностью предложит псевдонаучный контент о COVID-19 по сравнению с другими типами данного контента. Можно предположить, что недавние усилия YouTube по минимизации распространения фейкового контента о COVID-19 оказались успешными (Papadamou et al., 2022: 723-734).

Исследование Kaiser & Rauchfleisch (2019) сфокусировано на взаимосвязи рекомендательной системы и материалов сексуального характера с участием несовершеннолетних. Проанализировав и структурировав рекомендательную систему, авторы определили потенциальные пути к подобному контенту. Установлено, что самый неприемлемый контент находится не сразу, а в десяти кликах по рекомендациям. Видео были случайно обнаружены при анализе бразильских политических видео (Kaiser & Rauchfleisch, 2019: 1-22).

В своём исследовании Stöcker & Preuss (2020) сообщают, что сексуальный контент не является целью рекомендательной системы, скорее, это побочный эффект. Так, система иногда приводит к продвижению неприемлемого контента на более широкую аудиторию, чаще при

активной функции автовоспроизведения видео (Stöcker & Preuss, 2020: 359-375), отключение которой, по мнению Alfano et al. (2021), может побудить пользователей задуматься о том, что они хотели бы посмотреть дальше, а не позволять принимать решения рекомендательной системе (Alfano et al., 2021: 835-858).

Таким образом, авторы исследований, указанных нами как «получившие смешанные результаты», согласны с доступностью вредоносного контента, но существуют разногласия по поводу вероятности доступа к неприемлемому контенту через рекомендательную систему.

### Заключение

В целом результаты показывают, что рекомендательная система YouTube может облегчить поиск проблемного контента, однако необходимы дальнейшие исследования, чтобы понять основные механизмы, в которых происходит этот процесс, и его последствия. Отметим, что основные действия по минимизации вредоносного контента должны быть предприняты сетью YouTube.

Следует указать, что 20 из 22 исследований, включённых в данный мета-анализ, не рассматривали персонализацию пользователей. Чаще рекомендации основывались на исходных видео и не учитывали общие факторы (к примеру, взаимодействие пользователей с другими продуктами Google и их личная история поиска). Многие работы в этой области используют алгоритмы случайного обхода для определения вероятности обнаружения проблемного контента через рекомендательную систему и, таким образом, не учитывают персонализацию пользователя. Только в двух исследованиях использовался плагин для отслеживания реальных пользователей YouTube и их рекомендаций (Chen et al., 2022; Hosseinmardi et al., 2020). В будущих исследованиях следует сосредоточиться на вопросе влияния персонализации пользователей на рекомендации YouTube, что может послужить полезной информацией о рекомендательной системе YouTube и путях распространения вредоносного контента.

Немаловажно то, что социальная сеть YouTube обладает превосходными технологиями и чётким пониманием того, как работает её рекомендательная система, и можно было бы добиться значительных успехов, если бы YouTube тесно сотрудничал с исследователями, однако в

настоящее время вопрос коллаборации остаётся нерешённым, хотя YouTube постоянно принимает меры по демонетизации вредоносного контента, сокращению вредоносной дезинформации и изменениям в алгоритме рекомендаций.

Таким образом, стремительное развитие социальных сетей ставит перед пользователями новые задачи. Пользователи должны обладать информационной грамотностью, влияющей на способность формирования собственной позиции и лидерских качеств (Касымбекова & Шынгысова, 2022: 69), и с осторожностью относиться к контенту, который они потребляют, в том числе в сети YouTube.

Информационная грамотность – это не только свободное выражение своего мнения, но и правильная сортировка информации, воспринимаемой из внешнего мира, критическое осмысление постов читающих людей в социальной сети, грамотное потребление информационного контента. Отсутствие важных компетенций в

работе с информацией приводит к распространению фейков и конфликтам (Kassymbekova & Shyngyssova, 2022: 69). Осведомлённость пользователей о потенциальном влиянии социальных сетей на их жизнь может привести к осознанному поведению, предотвращающему просмотр вредоносного контента.

Со стороны YouTube также может быть принят предупредительный ряд мер. Такие изменения в работе рекомендательной системы, как, к примеру, размещение сообщений об опасности экстремизма в виде рекламы в видеороликах, определённых системой как вредоносные, могут оградить пользователей от проблемного контента. Целесообразно введение ботов, способных критически оценивать и опровергать возможную дезинформацию, выявляя логические ошибки или риторические недостатки в вводящих в заблуждение видеоматериалах до того, как их просмотрят пользователи.

## Литература

- Абдрашев, Р.М. (2016). Противодействие интернет-пропаганде экстремизма в Республике Казахстан. *Вестник Сибирского юридического института МВД России*, 1 (22): 58-62.
- Abul-Fottouh, D., Song, M.Y., & Gruzd, A. (2020). Examining algorithmic biases in YouTube's recommendations of vaccine videos. *International Journal of Medical Informatics*, 140: 104175. <https://doi.org/10.1016/j.ijmedinf.2020.104175>
- Alfano, M., Fard, A.E., Carter, J.A., Clutton, P., & Klein, C. (2021). Technologically scaffolded atypical cognition: The case of YouTube's recommender system. // *Synthese*, 199 (1): 835-858.
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110 (3): 629-676. <https://doi.org/10.1257/aer.20190658>
- AVAAZ (2020). Why is YouTube broadcasting climate misinformation to millions? [Report]. *Avaaz*. [https://secure.avaaz.org/campaign/en/youtube\\_climate\\_misinformation/](https://secure.avaaz.org/campaign/en/youtube_climate_misinformation/)
- Chen, A., Nyhan, B., Reifler, J., Robertson, R., & Wilson, C. (2022). Exposure to Alternative & Extremist Content on YouTube [Report]. *Anti-Defamation League*. <https://www.adl.org/resources/reports/exposure-to-alternative-extremist-content-on-youtube>
- Courtois, C., & Timmermans, E. (2018). Cracking the Tinder Code: An Experience Sampling Approach to the Dynamics and Impact of Platform Governing Algorithms. *Journal of Computer Mediated Communication*, 23 (1): 1-16. <https://doi.org/10.1093/jcmc/zmx001>
- Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos. *ArXiv preprint arXiv: 2003.03318*. <https://doi.org/10.48550/ARXiv.2003.03318>
- Fano A. N. et al. (2022). Evaluation of YouTube as a Source of Information Regarding Syndactyly. *Pediatrics*, 149 (1): 779.
- Green, S.J. (2019). God told me he was a lizard': Seattle man accused of killing his brother with a sword. *The Seattle Times*. <https://www.seattletimes.com/seattle-news/crime/god-told-me-he-was-alizard-seattle-man-accused-of-killing-his-brother-with-a-sword/>
- Hosseinmardi, H. et al. (2020). Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube. *ArXiv preprint arXiv: 2011.12843*. <https://doi.org/10.48550/arXiv.2011.12843>
- Hussein, E., Juneja, P., & Mitra, T. (2020). Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW1): 1-27. <https://doi.org/10.1145/3392854>
- Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, 78, 90-97. <https://doi.org/10.1016/j.chb.2017.09.022>
- Kaiser, J., & Rauchfleisch A. (2020). Birds of a feather get recommended together: Algorithmic homophily in YouTube's channel recommendations in the United States and Germany. *Social Media + Society*, 6 (4): 2056305120969914. <https://doi.org/10.1177/2056305120969914>
- Kaiser, J., & Rauchfleisch, A. (2019). The implications of venturing down the rabbit hole. *Internet Policy Review*, 8 (2): 1-22.
- Kaiser, J., Rauchfleisch, A., & Cordova, Y. (2021). Comparative Approaches to Mis / Disinformation| Fighting Zika with Honey: An Analysis of YouTube's Video Recommendations on Brazilian YouTube. *International Journal of Communication*, 15: 1244-1262.

- Касымбекова, Н.М., Шынгысова, Н.Т. (2022). Роль социальных сетей в формировании общественного мнения. *Вестник КазНУ им. Аль-Фараби. Серия Журналистики*, 3 (65): 69. <https://doi.org/10.26577/HJ.2022.v65.i3.07>
- Ledwich, M., & Zaitsev, A. (2019). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *ArXiv preprint arXiv: 1912.11211*. <https://doi.org/10.48550/arXiv.1912.11211>
- Ledwich, M., Zaitsev, A., & Laukemper, A. (2022). Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations. *First Monday*, 27 (12). <https://doi.org/10.5210/fm.v27i12.12552>
- Müller, K., & Schwarz, C. (2021). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*, 19 (4): 2131–2167. <https://doi.org/10.1093/jeea/jvaa045>
- Munger, K., & Phillips, J. (2022). Right-wing YouTube: A supply and demand perspective. *The International Journal of Press / Politics*, 27 (1): 186-219. <https://doi.org/10.1177/194016122096476>
- Nickles, M.A., Rustad, A.M., Ogbuefi, N., McKenney, J.E., & Stout, M. (2022). What's being recommended to patients on social media? A cross-sectional analysis of acne treatments on YouTube. *Journal of the American Academy of Dermatology*, 86 (4): 920-923. <https://doi.org/10.1016/j.jaad.2021.03.053>
- Nienierza, A., Reinemann, C., Fawzi, N., Riesmeyer, C., & Neumann, K. (2021). Too dark to see? Explaining adolescents' contact with online extremism and their ability to recognize it. *Information, Communication & Society*, 24 (9): 1229-1246. <https://doi.org/10.1080/1369118X.2019.1697339>
- Papadamou, K. et al. (2020). Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children. *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 522-533. <https://doi.org/10.1609/icwsm.v14i1.7320>
- Papadamou, K. et al. (2021). "How over is it?" Understanding the Incel Community on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2): 1-25. <https://doi.org/10.1145/3479556>
- Papadamou, K. et al. (2022). "It is just a flu": Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations. *Proceedings of the International AAAI Conference on Web and Social Media*, 16: 723-734. <https://doi.org/10.1609/icwsm.v16i1.19329>
- Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A.F., & Meira, W. (2020). Auditing Radicalization Pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 131-141. <https://doi.org/10.1145/3351095.3372879>
- Röchert, D., Weitzel, M., & Ross, B. (2020). The homogeneity of right-wing populist and radical content in YouTube recommendations. *International Conference on Social Media and Society*: 245-254. <https://doi.org/10.1145/3400806.3400835>
- Roose, K. (2019). The making of a YouTube radical. *The New York Times*, 8. <https://rhet104.commacafe.org/wp-content/uploads/2021/05/Making-of-a-YouTube-Radical.pdf>
- Schaub, M., & Morisi, D. (2020). Voter mobilisation in the echo chamber: Broadband internet and the rise of populism in Europe. *European Journal of Political Research*, 59 (4): 752-773. <https://doi.org/10.1111/1475-6765.12373>
- Schmitt, J.B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube: Recommendation algorithms. *Journal of Communication*, 68 (4): 780-808. <https://doi.org/10.1093/joc/jqy029>
- Spinelli, L., & Crovella, M. (2020). How YouTube leads privacy-seeking users away from reliable information. *Adjunct publication of the 28th ACM conference on user modeling, adaptation and personalization*: 244-251. <https://doi.org/10.1145/3386392.3399566>
- Stöcker, C., & Preuss, M. (2020). Riding the wave of misclassification: how we end up with extreme YouTube content. *International Conference on Human-Computer Interaction*: 359-375. [https://doi.org/10.1007/978-3-030-49570-1\\_25](https://doi.org/10.1007/978-3-030-49570-1_25)
- Williams, M.L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, 60 (1): 93-117. <https://doi.org/10.1093/bjc/azz049>
- Есенбекова, Ү.М., Алдабергенова, Ж.Ж., Маманкул, А.А., Смаилова, Б.А., Толегенова, С.Т. (2022). Элеуметтік медианың жастар белсенділігіне әсері. *Вестник КазНУ им. Аль-Фараби. Серия Журналистики*, 63 (1): 53. <https://doi.org/10.26577/HJ.2022.v63.i1.06>

## References

- Abdrashev, R.M. (2016). Protivodejstvie internet-propagande jekstremizma v Respublike Kazahstan [Counteraction to extremism propaganda on the Internet in the Republic of Kazakhstan]. *Bulletin of the Siberian Law Institute of the Ministry of Internal Affairs of Russia*, vol. 1, no 22, pp. 58–62.
- Abul-Fottouh, D., Song, M.Y., & Grudz, A. (2020). Examining algorithmic biases in YouTube's recommendations of vaccine videos. *International Journal of Medical Informatics*, 140: 104175. <https://doi.org/10.1016/j.ijmedinf.2020.104175>
- Alfano, M., Fard, A.E., Carter, J.A., Clutton, P., & Klein, C. (2021). Technologically scaffolded atypical cognition: The case of YouTube's recommender system. // *Synthese*, 199 (1): 835-858.
- Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110 (3): 629-676. <https://doi.org/10.1257/aer.20190658>
- AVAAZ (2020). Why is YouTube broadcasting climate misinformation to millions? [Report]. *Avaaz*. [https://secure.avaaz.org/campaign/en/youtube\\_climate\\_misinformation/](https://secure.avaaz.org/campaign/en/youtube_climate_misinformation/)

- Chen, A., Nyhan, B., Reifler, J., Robertson, R., & Wilson, C. (2022). Exposure to Alternative & Extremist Content on YouTube [Report]. *Anti-Defamation League*. <https://www.adl.org/resources/reports/exposure-to-alternative-extremist-content-on-youtube>
- Courtois, C., & Timmermans, E. (2018). Cracking the Tinder Code: An Experience Sampling Approach to the Dynamics and Impact of Platform Governing Algorithms. *Journal of Computer Mediated Communication*, 23 (1): 1-16. <https://doi.org/10.1093/jcmc/zmx001>
- Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos. *ArXiv preprint arXiv: 2003.03318*. <https://doi.org/10.48550/ARXiv.2003.03318>
- Fano A. N. et al. (2022). Evaluation of YouTube as a Source of Information Regarding Syndactyly. *Pediatrics*, 149 (1): 779.
- Green, S.J. (2019). God told me he was a lizard': Seattle man accused of killing his brother with a sword. *The Seattle Times*. <https://www.seattletimes.com/seattle-news/crime/god-told-me-he-was-alizard-seattle-man-accused-of-killing-his-brother-with-a-sword/>
- Hosseinmardi, H. et al. (2020). Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube. *ArXiv preprint arXiv: 2011.12843*. <https://doi.org/10.48550/arXiv.2011.12843>
- Hussein, E., Juneja, P., & Mitra, T. (2020). Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW1): 1-27. <https://doi.org/10.1145/3392854>
- Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, 78, 90-97. <https://doi.org/10.1016/j.chb.2017.09.022>
- Kaiser, J., & Rauchfleisch A. (2020). Birds of a feather get recommended together: Algorithmic homophily in YouTube's channel recommendations in the United States and Germany. *Social Media + Society*, 6 (4): 2056305120969914. <https://doi.org/10.1177/2056305120969914>
- Kaiser, J., & Rauchfleisch, A. (2019). The implications of venturing down the rabbit hole. *Internet Policy Review*, 8 (2): 1-22.
- Kaiser, J., Rauchfleisch, A., & Cordova, Y. (2021). Comparative Approaches to Mis / Disinformation| Fighting Zika with Honey: An Analysis of YouTube's Video Recommendations on Brazilian YouTube. *International Journal of Communication*, 15: 1244-1262.
- Kassymbekova, N.M. & Shyngyssova, N.T. (2022). Rol' social'nyh setej v formirovanii obshhestvennogo mnenija [The role of social networks in public opinion shaping]. *Herald Of Journalism*, vol. 3, no 65, p. 69. <https://doi.org/10.26577/HJ.2022.v65.i3.07>
- Ledwich, M., & Zaitsev, A. (2019). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *ArXiv preprint arXiv: 1912.11211*. <https://doi.org/10.48550/arXiv.1912.11211>
- Ledwich, M., Zaitsev, A., & Laukemper, A. (2022). Radical bubbles on YouTube? Revisiting algorithmic extremism with personalised recommendations. *First Monday*, 27 (12). <https://doi.org/10.5210/fm.v27i12.12552>
- Müller, K., & Schwarz, C. (2021). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*, 19 (4): 2131–2167. <https://doi.org/10.1093/jeea/jvaa045>
- Munger, K., & Phillips, J. (2022). Right-wing YouTube: A supply and demand perspective. *The International Journal of Press / Politics*, 27 (1): 186-219. <https://doi.org/10.1177/194016122096476>
- Nickles, M.A., Rustad, A.M., Ogbuefi, N., McKenney, J.E., & Stout, M. (2022). What's being recommended to patients on social media? A cross-sectional analysis of acne treatments on YouTube. *Journal of the American Academy of Dermatology*, 86 (4): 920-923. <https://doi.org/10.1016/j.jaad.2021.03.053>
- Nienierza, A., Reinemann, C., Fawzi, N., Riesmeyer, C., & Neumann, K. (2021). Too dark to see? Explaining adolescents' contact with online extremism and their ability to recognize it. *Information, Communication & Society*, 24 (9): 1229-1246. <https://doi.org/10.1080/1369118X.2019.1697339>
- Papadamou, K. et al. (2020). Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children. *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 522-533. <https://doi.org/10.1609/icwsm.v14i1.7320>
- Papadamou, K. et al. (2021). "How over is it?" Understanding the Incel Community on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2): 1-25. <https://doi.org/10.1145/3479556>
- Papadamou, K. et al. (2022). "It is just a flu": Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations. *Proceedings of the International AAAI Conference on Web and Social Media*, 16: 723-734. <https://doi.org/10.1609/icwsm.v16i1.19329>
- Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A.F., & Meira, W. (2020). Auditing Radicalization Pathways on YouTube. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 131-141. <https://doi.org/10.1145/3351095.3372879>
- Röcher, D., Weitzel, M., & Ross, B. (2020). The homogeneity of right-wing populist and radical content in YouTube recommendations. *International Conference on Social Media and Society*: 245-254. <https://doi.org/10.1145/3400806.3400835>
- Roose, K. (2019). The making of a YouTube radical. *The New York Times*, 8. <https://rhet104.commacafe.org/wp-content/uploads/2021/05/Making-of-a-YouTube-Radical.pdf>
- Schaub, M., & Morisi, D. (2020). Voter mobilisation in the echo chamber: Broadband internet and the rise of populism in Europe. *European Journal of Political Research*, 59 (4): 752-773. <https://doi.org/10.1111/1475-6765.12373>
- Schmitt, J.B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube: Recommendation algorithms. *Journal of Communication*, 68 (4): 780-808. <https://doi.org/10.1093/joc/jqy029>
- Spinelli, L., & Crovella, M. (2020). How YouTube leads privacy-seeking users away from reliable information. *Adjunct publication of the 28th ACM conference on user modeling, adaptation and personalization*: 244-251. <https://doi.org/10.1145/3386392.3399566>

Stöcker, C., & Preuss, M. (2020). Riding the wave of misclassification: how we end up with extreme YouTube content. *International Conference on Human-Computer Interaction*: 359-375. [https://doi.org/10.1007/978-3-030-49570-1\\_25](https://doi.org/10.1007/978-3-030-49570-1_25)

Williams, M.L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, 60 (1): 93-117. <https://doi.org/10.1093/bjc/azz049>

Yessenbekova, U., Aldabergenova, Z., Mamankul, A., Smailova, B., & Tolegenova, S. (2022). Äleumettik medianyň jastar belsendiligine äseri [The impact of social networks on youth activity]. *Herald Of Journalism*, vol. 63, no 1, p. 53. <https://doi.org/10.26577/HJ.2022.v63.i1.06>