



ПОЛИТИЧЕСКАЯ СОЦИОЛОГИЯ/POLITICAL SOCIOLOGY

DOI: <https://doi.org/10.60797/IRJ.2026.167.114> EDN: GFIDNV

ОСОБЕННОСТИ И ПЕРСПЕКТИВЫ РАСПРЕДЕЛЕННОЙ МОДЕРАЦИИ КОНТЕНТА

Научная статья

Манжула В.Г.^{1,*}, Манжула Е.В.²¹ORCID : 0000-0002-0672-9092;¹ Редакция газеты «Вечерняя Москва», Москва, Российская Федерация² Военный университет имени князя Александра Невского, Москва, Российская Федерация

* Корреспондирующий автор (manjula[at]bk.ru)

Предложена: 13.03.2026; Принята: 23.04.2026; Опубликовано: 18.05.2026

Аннотация

В статье рассмотрены особенности и механизмы распределенной (децентрализованной) модерации контента различных интернет-сообществ. Децентрализованная модерация обладает рядом достоинств, но имеет и ряд недостатков. Показано, что онлайн-коммуникация, как и любая другая, наиболее эффективно служит и удовлетворяет своих участников, если она происходит в духе уважения и сотрудничества, при этом все участвующие стороны в идеале разделяют общую цель. Децентрализованная модерация контента стимулирует онлайн-сообщества размышлять, обсуждать и согласовывать цели своего сообщества, что может привить как участникам, так и модераторам культуру добродетельной ответственности. Подчеркивается, что распределенная модерация наиболее подходит для проектирования систем с приоритетом на создание здоровых дискуссионных пространств.

Ключевые слова: модератор, контент, децентрализация, интернет-сообщество, сотрудничество.

FEATURES AND PROSPECTS OF DISTRIBUTED CONTENT MODERATION

Research article

Manjula V.G.^{1,*}, Manzhula Y.V.²¹ORCID : 0000-0002-0672-9092;¹ Newspaper editorial office «Evening Moscow», Moscow, Russian Federation² Prince Alexander Nevsky Military University, Moscow, Russian Federation

* Corresponding author (manjula[at]bk.ru)

Suggested: 13.03.2026; Accepted: 23.04.2026; Published: 18.05.2026

Abstract

The article discusses the features and mechanisms of distributed (decentralized) content moderation of various online communities. Decentralized moderation has a number of advantages, but it also has a number of disadvantages. It has been shown that online communication, like any other, most effectively serves and satisfies its participants if it takes place in a spirit of respect and cooperation, while all parties involved ideally share a common goal. Decentralized content moderation encourages online communities to reflect, discuss, and coordinate their community's goals, which can instill a culture of virtuous responsibility in both participants and moderators. It is emphasized that distributed moderation is most suitable for designing systems with a priority on creating healthy discussion spaces.

Keywords: moderator, content, decentralization, online community, collaboration.**Введение**

Развитие информационных технологий принесло с собой достаточно новых проблем, связанных как с манипулированием общественным мнением [1], так и с распространением ложной информации [2], [3]. Модерация интернет-контента во много позволяет избежать распространение как ложной информации, так и откровенной дезинформации. Исследования последних лет значительно расширили общепринятые представления о модерации онлайн-контента [4]. В настоящее время общепризнано, что модерация контента осуществляется на разных уровнях технологического процесса; нет жесткой границы между модерацией контента, понимаемой как удаление контента, и более мягкими формами модерации, такими как:

– продвижение или торможение контента;

– модерация контента часто предполагает сложное взаимодействие различных участников и видов деятельности.

Несмотря на это, по своей сути, модерацию контента по-прежнему можно определить как проверку, оценку, категоризацию, одобрение или удаление контента в соответствии с соответствующими коммуникационными и издательскими политиками [5].

В рамках этой концепции распределенная (децентрализованная) модерация контента предполагает и поддерживает инфраструктуру модерации, в которой полномочия и ответственность за модерацию распределены между множеством участников или организаций. Важно отметить, что децентрализация не является бинарной противоположностью централизованной модерации, а лишь указывает на относительное распределение полномочий по принятию решений о распространении контента.

История и развитие

Учитывая, что изначально онлайн-общение осуществлялось в основном с помощью децентрализованных протоколов (например, SMTP или TTP), а не корпоративных веб-сайтов, неудивительно, что на заре развития Интернета модерация контента также была в значительной степени децентрализована. Наиболее показательный пример такой ранней децентрализованной практики модерации контента представляет собой Usenet, распределенная система многопоточных форумов обмена сообщениями (так называемых групп новостей), распространяемая сетью независимых серверов. В то время как в самом Usenet не было универсальной политики в отношении контента или центрального органа, который мог бы обеспечить ее соблюдение, создатели групп новостей могли принять решение о «модерировании» группы. При этом каждая новая публикация пользователя требовала предварительного одобрения модератором, прежде чем она будет распространена среди других пользователей. Однако эта система предварительного одобрения стала узким местом и критической точкой отказа, от которой большинство создателей предпочли отказаться. Изначально отсутствие в Usenet эффективных инструментов модерации оставалось в значительной степени несущественным, так как большинство его первых пользователей принадлежали к однородной группе профессионалов, у которых были серьезные стимулы соблюдать общий сетевой этикет, чтобы не подвергать опасности свою репутацию в группе.

Однако, как только все более разнообразные демографические группы получили частный доступ в Интернет, такие дискурсивные соглашения было трудно поддерживать. Спам стал еще одной огромной проблемой, серьезно ставящей под угрозу удобство использования сервиса. Возможность серверов Usenet блокировать отдельных пользователей, размещать только новостные группы с определенными учетными данными или не распространять контент с других серверов, которые, как считается, содержат чрезмерное количество спама или другого «плохого» контента, предоставляла лишь довольно примитивные инструменты для решения этих проблем. Не в последнюю очередь и по этой причине Usenet, хотя и продолжает существовать, постепенно теряет свою актуальность.

В эпоху Web 2.0 появились гораздо более централизованные дискуссионные площадки, которые до сих пор считаются доминирующими социальными сетями, такими как Twitter* (теперь называется X), Facebook**, Instagram** или TikTok. Большинство из этих платформ внедрили единую политику в отношении контента в масштабах всей платформы, которая обеспечивается, в основном, централизованными механизмами модерации [6], [7]. Однако даже в эпоху «закрытых экосистем» некоторые сервисы сохранили децентрализованные формы управления контентом, ориентированные на сообщество. Reddit, где пользователи обсуждают темы на форумах по интересам (так называемых сабреддитах), пожалуй, является наиболее значимым примером такой децентрализованной модерации контента [8]. Хотя Reddit установил (возможно, довольно расплывчатую) общесайтовую политику в отношении контента, которую его так называемые администраторы могут обеспечивать, например, удаляя контент, блокируя учетные записи или удаляя целые сабреддиты, львиная доля модерации выполняется собственными модераторами сабреддитов. Обычно модераторами являются либо пользователи, создавшие сабреддит, либо другие пользователи, назначенные этими создателями. Модераторы должны не только соблюдать правила, действующие на сайте Reddit, но и устанавливать и обеспечивать соблюдение правил, специфичных для сабреддита.

Стремясь сделать обсуждение максимально увлекательным и соответствующим теме, эти правила зачастую выходят далеко за рамки запрета «вредного» контента и подробно предписывают, какие посты и комментарии разрешены [9]. Поскольку модерацией сабреддитов часто занимаются сами участники обсуждения, эта форма модерации, управляемая пользователями, иногда также называется (общественной) самомодерацией. Другой, хотя и более ограниченный, пример общественной модерации — функция «Community notes» на платформе X*. «Community notes», предназначенные для краудсорсинговой проверки фактов, позволяют пользователям добавлять дополнительную информацию к публикациям других пользователей (например, заметку, подтверждающую недостоверность исходного сообщения), если достаточное количество пользователей с «разными взглядами» оценили дополнительную информацию как полезную. Другие, в основном централизованные платформы, также недавно начали частично полагаться на краудсорсинг модерации, например, путем введения «доверенных модераторов» (YouTube, Facebook*, TikTok) или групповая самомодерация (Discord), возможно, иногда с целью сокращения расходов на «доверие и безопасность» и передачи сложных решений по модерации на аутсорсинг [10].

Наконец, в попытке предложить пользователям сети, альтернативы доминированию крупных социальных сетей, в последнее время возобновился интерес и началось интенсивное развитие по-настоящему децентрализованных онлайн-коммуникационных инфраструктур, возглавляемых протоколом ActivityPub и его флагманским сервисом социальных сетей Mastodon. Подобно Usenet, Mastodon не представляет собой единый веб-сайт, а скорее состоит из множества объединенных в сеть (федеративных) независимых серверов (так называемых «instances»). Каждый «instances» может самостоятельно определять свои локальные правила. Он также может блокировать другие «instances», делая контент, размещенный на таких «instances», невидимым для своих пользователей. Однако не существует центральной организации высшего уровня, способной запретить или удалить отдельные «instances» или ввести политику в отношении контента, выходящего за рамки «instances». Как и в случае с другими упомянутыми примерами распределенной модерации контента, пользователи, как правило, не могут напрямую влиять на политику в отношении контента «instances». Более того, в июле 2023 года Meta запустила Threads**, сервис микроблогинга, вдохновленный Twitter*, который, как она обещала, вскоре будет совместим с Activity Pub. В этом случае, Threads** представлял бы собой настоящий гибрид: экземпляр в децентрализованной сети, который сам по себе демонстрирует сильно централизованную форму управления контентом.

Очень похожим примером является Bluesky, еще один новый сервис микроблогов, разработанный на основе нового децентрализованного сетевого протокола под названием AT-protocol. Помимо децентрализации модерации, обеспечиваемой федеративной архитектурой протокола, Bluesky также планировал предоставить пользователям больше возможностей для индивидуальной модерации с помощью таких функций, как настраиваемые механизмы



выдачи рекомендаций и другое «middleware» для модерации. В качестве микропримера децентрализованной модерации контента можно рассмотреть очень популярные в России чаты жильцов, проживающих в одном доме. Такие чаты организуются в мессенджерах типа Max, Telegram, WhatsApp** и т.п. Модераторами таких чатов, как правило, являются неравнодушные жители этих домов, которые следят не только за выполнением требований законодательства, но и за тематикой и содержанием контента, размещаемого в чатах. Причем в одном доме может быть несколько чатов, тематически отличающихся друг от друга (животные, проезд во двор, капитальный ремонт дома и т.д.) и правила модерации могут в них отличаться.

Не оставим без внимания и российские социальные сети. В 2026 году на российском рынке соцмедиа продолжается этап перераспределения. После ухода одних игроков и роста других сформировалась устойчивая группа [11]. В настоящее время наиболее популярными, по оценке Mediascope, являются следующие ресурсы: ВКонтакте, Telegram, Max, YouTube, Дзен, TikTok, Одноклассники, Rutube, VK Видео и др. Рассмотрим возможности и используемые технологии модерации контента в ведущих российских социальных сетях. Так в соцсети ВКонтакте существуют общие требования к тому, что нельзя писать и публиковать в ВК. Это относится и к сообществам, и к аккаунтам обычных пользователей. Исполнение требований по контенту пользователей обеспечивается модераторами централизованно, причем как в автоматизированном, так и в ручном режиме [12].

Модерация контента в VK Видео происходит централизованно и автоматически: после загрузки ролик проверяется и, если нарушений не выявлено, публикация происходит в течение нескольких минут [13]. В соцсети Одноклассники используется как централизованная, так и распределенная модерация контента:

- модератор может одобрить запрос и принять в группу или же отказать, также как и администратор группы, модератор может открывать новые темы и добавлять новые альбомы, также он может изменять созданные им альбомы и темы и удалять;

- модератор имеет права удалить находящиеся во всех альбомах фото и картинки, а также удалять сообщения, которые противоречат правилам группы и сайта Одноклассники в целом;

- модератор также наделен таким правом, как блокировать пользователей группы за некорректное поведение и нарушение правил группы и сайта [14].

Сводная информация о типах модерации контента в некоторых популярных соцсетях и мессенджерах приведены в таблице 1.

Таблица 1 - Типы модерации контента в соцсетях и мессенджерах

DOI: <https://doi.org/10.60797/IRJ.2026.167.114.1>

Название ресурса	Используемый тип модерации	
	Распределенный	Централизованный
Usenet	+	–
Twitter* (теперь называется X)	+	+
Facebook**	+	+
Instagram**	+	+
TikTok	+	+
Reddit	+	–
YouTube	+	+
Discord	+	–
Mastodon	+	–
Threads**	+	+
Bluesky	+	+
ВКонтакте	+	–
Telegram	+	–
Max	+	–
Дзен	+	+
WhatsApp**	+	–
Одноклассники	+	+
Rutube	–	+
VK Видео	–	+
Guns.ru Talks	+	–

* – соцсеть Twitter (ныне переименованная в X) заблокирована на территории России по требованию Генпрокуратуры.

** – принадлежит Meta, признана судом экстремистской и запрещена в Российской Федерации.

Типология

Как уже упоминалось, децентрализация как социотехнический процесс не может быть понята в бинарных терминах. Скорее, ее необходимо выражать в терминах степени распределения системы или инфраструктуры между различными «узлами» принятия решений. Кроме того, децентрализованная модерация контента может включать в себя множество практик и принимать различные формы [15]. Несмотря на эту неоднородность, можно типизировать различные формы децентрализованной модерации контента. В частности, можно выделить три различных фактора, способствующих децентрализации системы модерации контента: распределение юрисдикции/полномочий между различными узлами, относительная автономия любого узла управления (по отношению к другим, особенно вышестоящим узлам управления), а также легкость выхода или миграции. Все три этих фактора оказывают решающее влияние на степень (де)централизации данной системы модерации контента.

Первый фактор связан с формальным аспектом юрисдикции, то есть с тем, какие и сколько узлов управления обладают правом принятия решений или полномочиями в отношении конкретного контента. (Слово «юрисдикция» используется здесь в неюридическом смысле, а для того, чтобы указать, обладает ли конкретный узел правом вынесения решения и полномочиями в отношении данного онлайн-контента.) Различные узлы управления контентом могут обладать либо исключительной, либо частично совпадающей юрисдикцией в отношении определенного контента. Частично совпадающая юрисдикция приводит к децентрализации права принятия решений в отношении контента между различными узлами. Например, крупные платформы социальных сетей, как правило, обладают исключительной универсальной юрисдикцией в отношении большей части пользовательского контента, что означает, что собственная команда модераторов компании является единственным узлом управления, который может удалять контент для всех пользователей. В то же время, команды модераторов компании могут разделять юрисдикцию в отношении контента, размещенного в группах (например, на Facebook**), с администраторами этих групп [16]. Аналогичный случай совпадения юрисдикций можно наблюдать на Reddit, где полномочия администраторов сабреддита по модерированию определенного сабреддита не исключают возможности одновременной работы собственной команды Reddit по обеспечению безопасности в отношении такого контента.

Во-вторых, различные варианты децентрализованной модерации контента можно охарактеризовать на основе автономии узлов модерации контента от соответствующих полномочий других, особенно иерархически вышестоящих узлов управления. Например, с одной стороны, можно сказать, что практика модерации на крупных платформах социальных сетей сильно децентрализована, поскольку решения распределяются между огромными группами модераторов, часто также с участием независимых подрядчиков. Однако, в конечном счете, хотя эти системы явно предполагают определенную степень многополярности или многоуровневого управления, специфика и жесткость правил внутренней модерации редко оставляют модераторам значительную самостоятельность в принятии решений, что приводит к эффективной централизации системы модерации. Еще одним важным фактором при оценке автономии модерации является то, насколько технологически закреплены полномочия по принятию решений. Например, хотя Reddit предоставляет модераторам сабреддитов значительные полномочия по модерации, эти «привилегии» могут быть легко отозваны, например, если Reddit решит изменить структуру сайта или в большей степени полагаться на централизованную модерацию контента. Напротив, структуры управления контентом, жестко заложенные в технологических протоколах, таких как платформы на основе Usenet или ActivityPub, предоставляют децентрализованным узлам модерации гораздо большую безопасность и автономию, гарантируя, что их методы модерации не будут подорваны или изменены сверху. Другими словами, они укрепляют автономию на инфраструктурном уровне.

В-третьих, последний фактор, определяющий эффективный уровень децентрализации модерации на платформе, — это возможность и относительная легкость, с которой участники могут покинуть платформу или перейти на новые форумы. Чем проще и дешевле администраторам форума или обычным пользователям создать новый форум или перейти на него, тем выше вероятность того, что платформа действительно децентрализуется и предложит пользователям несколько площадок для обсуждения. Например, ни Twitter* (теперь «X»), ни Instagram**, ни TikTok не предоставляют пользователям возможности выбора между различными форумами/системами модерации, что исключает один из самых фундаментальных уровней потенциальной децентрализации [17], [18]. Reddit или Mastodon, с другой стороны, позволяют пользователям создавать новые сабреддиты или переходить на новые платформы. Разница между этими двумя последними примерами может заключаться в относительной стоимости миграции. Факторы, определяющие эту стоимость, включают техническую легкость, с которой пользователи могут перейти на другой форум, или риск потери результатов своей прошлой активности на платформе, таких как существующие подписчики или другие учетные данные.

* – соцсеть Twitter (ныне переименованная в X) заблокирована на территории России по требованию Генпрокуратуры.

** – принадлежит Meta, признана судом экстремистской и запрещена в Российской Федерации.

Достоинства и недостатки

Как централизованные, так и децентрализованные формы модерации контента обладают своими сильными и слабыми сторонами. Пожалуй, самым большим преимуществом децентрализованной модерации является возможность распределения обязанностей по модерации между теми, кто наиболее хорошо осведомлен о тонкостях конкретных онлайн-сообществ. Таким образом, модераторы часто глубоко заинтересованы в укреплении, процветании и достижении целей этих сообществ и приобретают конкретный опыт и ноу-хау в отношении того, как достичь этих целей. Децентрализованная модерация контента, основанная на принципах традиционного федерализма как средство ограничения власти, способствует системной эффективности, плюрализму, профессионализму, свободе личности и является потенциальным ответом на многие проблемы, связанные с распространенными в настоящее время моделями



экономики внимания, ориентированной на рекламу. Распределенная самоModerация сообществ также может привести к более демократичным формам онлайн-управления. Наконец, децентрализованная модерация контента может обеспечить дискурсивное разнообразие во всей сети, а также смягчить возникновение чрезмерной власти модераторов или «единых точек отказа» [19].

Децентрализованная модерация контента, однако, также сопряжена с рисками. Во-первых, децентрализованная модерация может создавать так называемые «эхо-камеры» и усиливать дискурсивную фрагментацию и поляризацию общества. Конечно, вопрос о том, действительно ли такие онлайн «эхо-камеры» нежелательны и можно ли их эффективно предотвращать, по-прежнему остается открытым. Однако другие есть и другие, более ощутимые недостатки децентрализованной модерации контента. Важнейшим из них, на наш взгляд, из них является то, что модерация требует больших ресурсов и времени. Децентрализованные методы модерации могут не только не поддаваться эффективному масштабированию, что довольно легко удается централизованным системам, но и могут страдать от отсутствия источников финансирования. Действительно, идеалы устойчивого самоModerирования сообщества могут противоречить логике получения прибыли, необходимой для доступа к традиционным источникам финансирования. Отсутствие источников финансирования и риск хронической перегрузки команд модераторов также означают, что дополнительные меры защиты пользователей, такие как повышение прозрачности или механизмы обжалования, иногда могут оказаться нежизнеспособными. Более того, некоторые согласованные атаки на онлайн-форумы, такие как скрытое проникновение ботов, могут быть обнаружены только системными администраторами; децентрализованным командам модерации часто не хватает данных и инструментов для противодействия таким действиям. Одним из решений этой проблемы может стать более широкое и доступное использование инструментов автоматической модерации, поскольку достижения в области технологий машинного обучения снижают затраты на обнаружение и удаление нежелательного контента. В любом случае, из-за скорости, масштаба и ресурсов, необходимых для такой модерации, риск коммерческого использования модерации контента в децентрализованных системах может привести к повторной централизации методов модерации [20].

Заключение

Вряд ли можно сомневаться в том, что онлайн-коммуникация, как и любая другая, наиболее эффективно служит и удовлетворяет своих участников, если она происходит в духе уважения и сотрудничества, при этом все участвующие стороны в идеале разделяют общую цель. Децентрализация модерации контента стимулирует онлайн-сообщества размышлять, обсуждать и согласовывать цели своего сообщества, что может привить как участникам, так и модераторам культуру коллективной ответственности. Модерация контента и управление сообществом также должны основываться на традиционных федералистских принципах субсидиарности и локализма, и особенно на том, что управление, как правило, должно осуществляться на самом низком уровне, если иное не оправдано (например, соображениями практичности) [21], [22]. Однако представленные примеры децентрализованной модерации также показывают, насколько важно, чтобы платформа или протокол предоставляли модераторам адекватные инструменты и социотехнические структуры для эффективного и результативного выполнения своих задач [23]. Более того, вышестоящие узлы управления, скорее всего, не смогут делегировать действия, связанные с предотвращением и смягчением структурных проблем и атак. Таким образом, подчеркивается необходимость проектирования систем с приоритетом на создание здоровых дискуссионных пространств. Это может означать отказ от приоритизации методов вовлечения и роста, которые противоречат интересам пользователей и здоровью сообществ. Это также подчеркивает важность возможностей миграции для пользователей, которые гарантируют, что модерация сообщества будет по-прежнему соответствовать потребностям сообщества [24].

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Божич В.И. Комплексная информационно-психологическая безопасность корпорации знаний в открытом информационном пространстве сетевой экономики: учебное пособие / В.И. Божич, В.Г. Манжула. — Москва : ИКТ, 2013. — 211 с.
2. Манжула В.Г. Понятие псевдоинформации в современном мировом информационном пространстве / В.Г. Манжула, Е.В. Манжула // Актуальные исследования. — 2026. — № 6-2(292). — С. 48–60.
3. Манжула В.Г. Дезинформация и мисинформация — две стороны одной медали / В.Г. Манжула, Е.В. Манжула // Актуальные исследования. — 2026. — № 9-1(295). — С. 92–95.
4. Zane L.B. Perceptions of e-moderators about their roles and functions in moderating electronic mailing lists / L.B. Zane, P.C. Mauri // Distance Education. — 2000. — № 1(21). — P. 81–100.
5. Binns R. Like trainer, like bot? Inheritance of bias in algorithmic content moderation / R. Binns, M. Veale, M. Van Kleek et al. // International Conference on Social Informatics. — 2017. — P. 405–415.



6. Buni C. The Trauma Floor: The secret lives of Facebook moderators in America / C. Buni. — URL: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona/> (accessed: 12.03.2026).
7. Buni C. The secret rules of the internet: The murky history of moderation, and how it's shaping the future of free speech / C. Buni. — 2016. — URL: <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech/> (accessed: 12.03.2026).
8. Reddit : Moderator Code of Conduct. — URL: <https://redditinc.com/> (accessed: 12.03.2026).
9. Chandrasekharan E. The internet's hidden rules: An empirical study of reddit norm violations at micro, meso, and macro scales / E. Chandrasekharan, M. Samory, S. Jhaver et al. // Proceedings of the ACM on Human-Computer Interaction 2. — CSCW. — 2018. — P. 32.
10. Berin S. Congressional republicans push bills that would block kids access to content for ideological reasons / S. Berin. — 2026. — URL: <https://www.techdirt.com/2026/03/09/congressional-republicans-push-bills-that-would-block-kids-access-to-content-for-ideological-reasons/> (accessed: 12.03.2026).
11. Сурова А. Социальные сети в России — полный список популярных площадок со статистикой по аудитории / А. Сурова // eLama. — 2026. — URL: <https://elama.ru/blog/socialnye-seti-v-rossii/> (дата обращения: 03.04.2026)
12. Усачева Ю. Правила и ограничения «ВКонтакте» / Ю. Усачева // Calltouch Blog. — 2023. — URL: <https://www.calltouch.ru/blog/pravila-i-ogranicheniya-vkontakte/> (дата обращения: 06.04.2026).
13. MoreLogin: Rutube vs VK Видео в 2025 году. — 2025. — URL: <https://w.morelogin.com/ru/blog/rutube-vs-vk-video> (дата обращения: 06.04.2026).
14. Антропов С. Одноклассники ищут модераторов контента. Какую зарплату предлагают? / С. Антропов // Kadrof.RU. — 2023. — URL: <https://www.kadrof.ru/news/65463> (дата обращения: 06.04.2026).
15. Deimorz. AutoModerator – a bot for automating straightforward reddit moderation tasks and improving upon the existing spam-filter / Deimorz // TheoryOfReddit. — 2012. — URL: https://www.reddit.com/r/TheoryOfReddit/comments/onl2u/automoderator_a_bot_for_automating/ (accessed: 12.03.2026).
16. Sandra E.G. Ex-Content Moderator Sues Facebook, Saying Violent Images Caused Her PTSD / E.G. Sandra // The New York Times. — 2018. — URL: <https://www.nytimes.com/2018/09/25/technology/facebook-moderator-job-ptsd-lawsuit.html/> (accessed: 12.03.2026).
17. Gerrard Y. Beyond the hashtag: Circumventing content moderation on social media / Y. Gerrard // New Media & Society. — 2018. — № 12. — P. 4492–4511.
18. Gillespie T. Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media / T. Gillespie. — Yale University Press, 2018.
19. Graber J. Designing Decentralized Moderation / J. Graber. — 2022. — URL: <https://jaygraber.medium.com/designing-decentralized-moderation-a76430a8eab> (accessed: 12.03.2026)
20. Lampe C. Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums / C. Lampe, P. Zube, J. Lee et al. // Government Information Quarterly. — 2014. — № 2(31). — P. 317–326.
21. Yishai B. Federalism, subsidiarity, and the role of local governments in an age of global multilevel governance / B. Yishai. — URL: <https://ir.lawnet.fordham.edu/ulj/vol37/iss2/1> (accessed: 12.03.2026).
22. Kadri T.E. Juridical discourse for platforms / T.E. Kadri // Harvard Law Review. — 2022. — Vol. 136, № 2. — P. 163–204.
23. Murphy P.P. 'InfoWars' main YouTube channel is two strikes away from being banned / P.P. Murphy // CNN. — 2018. — URL: <https://edition.cnn.com/2018/02/23/us/infowars-youtube-videos-trnd/index.html> (accessed: 12.03.2026).
24. Gillespie T. Regulation of and by platforms / T. Gillespie // The SAGE Handbook of Social Media / ed. by J. Burgess, A. Marwick, T. Poell. — London : SAGE, 2018. — P. 254–278.

Список литературы на английском языке / References in English

1. Bozhich V.I. Kompleksnaya informatsionno-psikhologicheskaya bezopasnost' korporatsii znaniy v otkrytom informatsionnom prostranstve setevoy ekonomiki [Comprehensive information and psychological security of the knowledge corporation in the open information space of the network economy] : textbook / V.I. Bozhich, V.G. Manzhula. — Moscow : ICT, 2013. — 211 p. [in Russian]
2. Manzhula V.G. Ponyatiye psevdoinformatsii v sovremennom mirovom informatsionnom prostranstve [The concept of pseudo-information in the modern world information space] / V.G. Manzhula, E.V. Manzhula // Aktual'nyye issledovaniya [Current Research]. — 2026. — № 6-2(292). — P. 48–60.
3. Manzhula V.G. Dezinformatsiya i misinformatsiya — dve storony odnoy medalii [Disinformation and misinformation — two sides of the same coin] / V.G. Manzhula, E.V. Manzhula // Aktual'nyye issledovaniya [Current Research]. — 2026. — № 9-1(295). — P. 92–95.
4. Zane L.B. Perceptions of e-moderators about their roles and functions in moderating electronic mailing lists / L.B. Zane, P.C. Mauri // Distance Education. — 2000. — № 1(21). — P. 81–100.
5. Binns R. Like trainer, like bot? Inheritance of bias in algorithmic content moderation / R. Binns, M. Veale, M. Van Kleek et al. // International Conference on Social Informatics. — 2017. — P. 405–415.
6. Buni C. The Trauma Floor: The secret lives of Facebook moderators in America / C. Buni. — URL: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona/> (accessed: 12.03.2026).



7. Buni C. The secret rules of the internet: The murky history of moderation, and how it's shaping the future of free speech / C. Buni. — 2016. — URL: <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech/> (accessed: 12.03.2026).
8. Reddit : Moderator Code of Conduct. — URL: <https://redditinc.com/> (accessed: 12.03.2026).
9. Chandrasekharan E. The internet's hidden rules: An empirical study of reddit norm violations at micro, meso, and macro scales / E. Chandrasekharan, M. Samory, S. Jhaver et al. // *Proceedings of the ACM on Human-Computer Interaction 2.* — CSCW. — 2018. — P. 32.
10. Berin S. Congressional republicans push bills that would block kids access to content for ideological reasons / S. Berin. — 2026. — URL: <https://www.techdirt.com/2026/03/09/congressional-republicans-push-bills-that-would-block-kids-access-to-content-for-ideological-reasons/> (accessed: 12.03.2026).
11. Surova A. Sotsial'nyye seti v Rossii — polnyy spisok populyarnykh ploshchadok so statistiko po auditorii [Social networks in Russia — a complete list of popular platforms with audience statistics] / A. Surova // eLama. — 2026. — URL: <https://elama.ru/blog/socialnye-seti-v-rossii/> (accessed: 03.04.2026). [in Russian]
12. Usacheva Yu. Pravila i ogranicheniya "VKontakte" [Rules and restrictions of "VKontakte"] / Yu. Usacheva // Calltouch Blog. — 2023. — URL: <https://www.calltouch.ru/blog/pravila-i-ogranicheniya-vkontakte/> (accessed: 06.04.2026). [in Russian]
13. MoreLogin: Rutube vs VK Video v 2025 godu [MoreLogin: Rutube vs VK Video in 2025]. — 2025. — URL: <https://w.morelogin.com/ru/blog/rutube-vs-vk-video> (accessed: 06.04.2026). [in Russian]
14. Antropov S. Odnoklassniki ishchut moderatorov kontenta. Kakuyu zarplatu predlagayut? [Odnoklassniki are looking for content moderators. What salary are they offering?] / S. Antropov // Kadrof.RU. — 2023. — URL: <https://www.kadrof.ru/news/65463> (accessed: 06.04.2026). [in Russian]
15. Deimorz. AutoModerator – a bot for automating straightforward reddit moderation tasks and improving upon the existing spam-filter / Deimorz // TheoryOfReddit. — 2012. — URL: https://www.reddit.com/r/TheoryOfReddit/comments/onl2u/automoderator_a_bot_for_automating/ (accessed: 12.03.2026).
16. Sandra E.G. Ex-Content Moderator Sues Facebook, Saying Violent Images Caused Her PTSD / E.G. Sandra // *The New York Times.* — 2018. — URL: <https://www.nytimes.com/2018/09/25/technology/facebook-moderator-job-ptsd-lawsuit.html/> (accessed: 12.03.2026).
17. Gerrard Y. Beyond the hashtag: Circumventing content moderation on social media / Y. Gerrard // *New Media & Society.* — 2018. — № 12. — P. 4492–4511.
18. Gillespie T. Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media / T. Gillespie. — Yale University Press, 2018.
19. Graber J. Designing Decentralized Moderation / J. Graber. — 2022. — URL: <https://jaygraber.medium.com/designing-decentralized-moderation-a76430a8eab> (accessed: 12.03.2026)
20. Lampe C. Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums / C. Lampe, P. Zube, J. Lee et al. // *Government Information Quarterly.* — 2014. — № 2(31). — P. 317–326.
21. Yishai B. Federalism, subsidiarity, and the role of local governments in an age of global multilevel governance / B. Yishai. — URL: <https://ir.lawnet.fordham.edu/ulj/vol37/iss2/1> (accessed: 12.03.2026).
22. Kadri T.E. Juridical discourse for platforms / T.E. Kadri // *Harvard Law Review.* — 2022. — Vol. 136, № 2. — P. 163–204.
23. Murphy P.P. 'InfoWars' main YouTube channel is two strikes away from being banned / P.P. Murphy // CNN. — 2018. — URL: <https://edition.cnn.com/2018/02/23/us/infowars-youtube-videos-trnd/index.html> (accessed: 12.03.2026).
24. Gillespie T. Regulation of and by platforms / T. Gillespie // *The SAGE Handbook of Social Media* / ed. by J. Burgess, A. Marwick, T. Poell. — London : SAGE, 2018. — P. 254–278.