

ЭССЕ

ПРАВОВЫЕ МЕХАНИЗМЫ ЦИФРОВОЙ МОДЕРАЦИИ КОНТЕНТА НА HR-TECH ПЛАТФОРМАХ

С. Е. Шабронова

ООО «Хэдхантер»
125047, Россия, Москва, ул. Вторая Брестская, 48

Аннотация

Эссе посвящено анализу правовых механизмов цифровой модерации контента на *HR-tech*-платформах как элемента современной инфраструктуры регулирования цифрового рынка труда. В условиях цифровизации платформы перестают быть исключительно техническими посредниками и всё чаще выполняют функции частного регулирования. В этом контексте особое значение приобретает модерация контента, осуществляемая в отношении вакансий, резюме и сопутствующих информационных материалов. Целью исследования являются выявление юридической природы цифровой модерации на *HR-tech*-платформах, определение ее места в системе правового регулирования и обоснование ее роли как инструмента превентивного правоприменения в частноправовой цифровой среде. С учетом доктрины анализируются положения российского законодательства, нормы о статусе информационного посредника, а также подходы зарубежного регулирования, включая *Digital Services Act* (ЕС) и *Online Safety Act* (Великобритания). Дополнительно рассматриваются пользовательские соглашения и правила модерации конкретных платформ, что позволяет выявить практические модели премодерации, постмодерации, алгоритмической фильтрации и процедур обжалования. В эссе показано, что закрепление в российском праве специальных обязанностей владельцев сервисов размещения объявлений трансформирует традиционную модель ответственности информационного посредника и объективно предполагает внедрение активных механизмов модерации пользовательского контента. Обосновано, что цифровая модерация на *HR-tech*-платформах выходит за рамки технической фильтрации и приобретает значение механизма первичной юридической квалификации фактических обстоятельств, влекущей правовые последствия для пользователей. Особое внимание уделено дуализму правовой природы вакансий, которые в зависимости от содержания могут подпадать под различные правовые режимы, а также анализу алгоритмической модерации с учетом отраслевой специфики трудовых отношений. Аргументируется, что принцип полной прозрачности модерационных решений не является универсальным и в ряде случаев может снижать эффективность защиты прав соискателей и устойчивость платформенной экосистемы. Делается вывод о допустимости модели ограниченной прозрачности при условии сохранения механизмов апелляции и коррекции ошибок. Полученные результаты могут быть использованы при разработке пользовательских соглашений, внутренних правил модерации и алгоритмических систем *HR-tech*-платформ.

Ключевые слова

цифровая модерация, сервисы размещения объявлений, информационный посредник, вакансии, алгоритмическая модерация, цифровое правоприменение, рынок труда

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имеет спонсорской поддержки.

Для цитирования

Шабронова, С. Е. (2025). Правовые механизмы цифровой модерации контента на HR-Tech платформах. *Цифровое право*, 6(4), 8–20. <https://doi.org/10.38044/2686-9136-2025-6-19>

Поступило: 28.06.25, принято в печать: 16.11.25, опубликовано: 30.12.2025

ESSAYS

LEGAL MECHANISMS FOR DIGITAL CONTENT MODERATION ON HR TECH PLATFORMS

Svetlana E. Shabronova

Headhunter LLC

48, Vtoraya Brestskaya St., Moscow, Russia, 125047

Abstract

This essay examines the legal mechanisms of digital content moderation on *HR tech* platforms as an integral element of contemporary regulation of the digital labor market. In the context of digital transformation, online platforms increasingly move beyond the role of purely technical intermediaries and assume functions of private regulation. Within this framework, content moderation becomes a key instrument shaping the legal environment of interactions between employers and job seekers. The purpose of this study is to identify the legal nature of digital moderation on *HR tech* platforms, to determine its place within the system of legal regulation, and to substantiate its role as a form of preventive law enforcement in a private digital environment. On a doctrinal basis, the essay examines relevant provisions of Russian law, including rules on the status and liability model of the “information intermediary,” and compares them with foreign regulatory approaches, notably the EU *Digital Services Act* and the United Kingdom’s *Online Safety Act*. The analysis is complemented by a review of platform user agreements and moderation policies, enabling the identification of operational models of pre-moderation, post-moderation, algorithmic filtering, and appeals procedures. The essay demonstrates that the introduction of specific statutory obligations for operators of online classified services transforms the traditional model of information intermediary liability and objectively requires the implementation of proactive content moderation mechanisms. It is argued that digital moderation on HR tech platforms cannot be reduced to technical filtering; rather, it constitutes a mechanism of primary legal qualification of factual circumstances, the results of which entail legal consequences for platform users. Special attention is paid to the dual legal nature of job vacancy postings, which may fall under different legal regimes depending on their content, and to algorithmic moderation adapted to the sectoral specifics of labor relations. The study further argues that the principle of full transparency of moderation decisions is not universal and, in certain contexts, may undermine the effectiveness of fraud prevention and the stability of platform ecosystems. The essay substantiates the permissibility of limited transparency models, provided that accessible appeal mechanisms and error-correction procedures are maintained. The findings of this research may be applied in the

development of user agreements, internal moderation policies, and algorithmic moderation systems of HR tech platforms.

Keywords

digital moderation, HR tech platforms, online classified services, information intermediary, job vacancies, algorithmic moderation, digital law enforcement, labor market

Conflict of interest The author declares no conflict of interest.

Financial disclosure The study has no sponsorship.

For citation Shabronova, S. E. (2025). Legal mechanisms for digital content moderation on HR Tech platforms. *Digital Law Journal*, 6(4), 8–20. <https://doi.org/10.38044/2686-9136-2025-6-19>

Submitted: 28 Jun. 2025, accepted: 16 Nov. 2025, published: 30 Dec. 2025

Введение

Развитие цифровых технологий трансформировало не только способы реализации прав, но и механизмы формирования юридически значимых фактов. В научной литературе отмечается, что право в цифровой среде все чаще приобретает превентивный характер, ориентированный на упреждение конфликтов до их возникновения (Altukhov & Kashkin, 2021, pp. 86–94; Khabrieva & Chernogor, 2018, pp. 85–102).

Особую роль в этой трансформации играют цифровые платформы, которые из инструментов посредничества превращаются в субъектов частного регулирования: они устанавливают правила взаимодействия пользователей, контролируют их соблюдение и применяют меры воздействия (Casas-Cortés et al., 2023).

В контексте *HR-tech*-платформ — сервисов, обеспечивающих взаимодействие работодателей и соискателей, — ключевым инструментом такого регулирования выступает модерация контента.

Законодательство РФ не содержит легального определения термина «модерация». В Федеральном законе № 149-ФЗ¹ (далее — цитируется по номеру) он используется единожды в контексте обязанностей владельцев социальных сетей по мониторингу и удалению запрещённого контента. Однако для сервисов размещения объявлений, к которым относятся *HR-tech*-платформы, модерация приобретает специфическое значение: именно через неё реализуется обязанность владельца сервиса «не допускать использование сервиса» для распространения противоправной информации (ст. 10.7 Закона № 149-ФЗ).

В научной дискуссии акцент нередко делается на рисках для свободы выражения мнений и недостаточной прозрачности алгоритмов. Однако в контексте *HR-tech*-платформ модерация имеет иное функциональное назначение: она направлена на обеспечение качества вакансий, защиту прав соискателей и формирование безопасного цифрового пространства на рынке труда.

¹ Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27 июля 2006 г. № 149-ФЗ. Собрание законодательства РФ, № 31, 2006 г., ст. 3448 (часть I).

Цель настоящего эссе — анализ правовых механизмов цифровой модерации контента на *HR-tech*-платформах в контексте российского законодательства. Основная гипотеза исследования заключается в том, что цифровая модерация на *HR-tech*-платформах представляет собой механизм первичной юридической квалификации фактов, выполняющий превентивную правоприменительную функцию, эффективность которой достигается через модель ограниченной прозрачности при сохранении механизмов апелляции.

HR-tech-платформы как сервисы размещения объявлений

В российском праве правовой статус *HR-tech*-платформ целесообразно рассматривать через призму Закона № 149-ФЗ. В частности, применима статья 10.7 указанного Закона, регулирующая обязанности владельцев сервисов размещения объявлений.

HR-tech-платформы предоставляют пользователям возможность размещения информации о предложении работы, откликов и сопутствующих сведений, что позволяет квалифицировать их именно как сервисы размещения объявлений. Такой подход соответствует функциональной природе данных платформ и позволяет избежать некорректного отождествления *HR-tech*-сервисов с социальными сетями или иными видами цифровых платформ.

Статья 10.7 Закона № 149-ФЗ возлагает на владельцев сервисов размещения объявлений обязанности по недопущению использования сервиса для совершения уголовно наказуемых деяний; предотвращению распространения экстремистских, террористических, порнографических, насильственных материалов, а также информации, содержащей нецензурную брань; пресечению распространения сведений, порочащих граждан или группы лиц. В данном контексте цифровая модерация выступает не факультативной функцией платформы, а инструментом реализации ее законодательно закрепленных обязанностей.

Эволюция конструкции информационного посредника применительно к владельцам сервисов объявлений

До появления ст. 10.7 Закона № 149-ФЗ владелец сервиса объявлений формально рассматривался как информационный посредник (ст. 1253.1 ГК РФ²), нес ограниченную ответственность при знании о нарушении и бездействии.

Статья 10.7 Закона № 149-ФЗ, введенная в 2022 г., изменяет эту конструкцию, поскольку прямо возлагает публично-правовые обязанности и не ограничивается моделью удаления после уведомления, что связывает ответственность не с фактом знания, а с функциональной ролью платформы.

Таким образом, статья 10.7 Закона № 149-ФЗ позволяет сделать вывод о том, что российский законодатель возлагает на владельцев сервисов размещения объявлений обязанность по обеспечению контроля за распространяемым пользователями контентом. Несмотря на отсутствие прямого указания на обязанность предварительной модерации, требование Закона не допускать использование сервиса для распространения противоправной информации предполагает необходимость выстраивания владельцем сервиса организационных и технических механизмов модерации пользовательского контента.

² Гражданский кодекс Российской Федерации. Часть четвертая от 18 декабря 2006 г. № 230-ФЗ. Собрание законодательства РФ, 2006 г., № 52, ст. 5496 (часть I).

Законодательство ЕС, в частности Регламент о цифровых услугах (*Digital Services Act*)³, прямо требует от платформ внедрять механизмы контроля за распространением незаконного и вредного контента и предусматривать процессы обработки жалоб и действий по исправлению нарушений со стороны платформы.

Аналогичные тенденции наблюдаются в Великобритании: Закон об онлайн-безопасности (*Online Safety Act*) вводит обязанности для онлайн-сервисов по предотвращению распространения незаконного и опасного контента, включая защиту детей и удаление запрещенных материалов⁴.

С точки зрения теории права цифровую модерацию можно рассматривать как механизм юридической квалификации фактических обстоятельств. В классической доктрине юридический факт определяется как жизненное обстоятельство, с которым норма права связывает возникновение, изменение или прекращение правоотношений (Krasavchikov, 1958, p. 5). В цифровой среде фиксация и оценка таких обстоятельств всё чаще осуществляются автоматизированными системами.

Размещение вакансии, выявление нарушения правил платформы, установление несоответствия содержания вакансии требованиям законодательства — все эти действия фиксируются в цифровой системе и приобретают значение юридически значимых фактов. Решение о блоке или ограничении доступа пользователя представляет собой правовое последствие, наступающее в результате квалификации соответствующего факта.

Таким образом, цифровая модерация на *HR-tech*-платформах выполняет функцию первичного правоприменения в частноправовой цифровой среде. Платформа не заменяет государственные органы, однако осуществляет предварительную оценку информации и предотвращает распространение контента, потенциально нарушающего права пользователей или требования законодательства.

Типы модерации и их роль

Правовые механизмы на *HR-tech*-платформах включают элементы как обязательной, так и добровольной модерации. Обязательная модерация обусловлена требованиями законодательства, в том числе обязанностями владельца сервиса размещения объявлений по ст. 10.7 Закона № 149-ФЗ, о которых написано выше. Добровольная модерация осуществляется на основании внутренних правил платформы и направлена на повышение качества сервиса и доверия пользователей.

Внутренние стандарты модерации позволяют платформам устанавливать требования к содержанию вакансий, что прямо предусмотрено законом. Такая практика соответствует концепции сочетания правового регулирования и саморегулирования в цифровой среде, широко обсуждаемой в научной литературе (Khabrieva & Chernogor, 2018, p. 97).

Современные платформы используют премодерацию (проверку контента перед публикацией), постмодерацию (публикацию с последующей проверкой) и гибридные модели, сочетающие оба подхода⁵. Выбор конкретной модели определяется балансом между требованиями

³ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 Oct. 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), arts. 14–17, 2022 O.J. (L 277) 1.

⁴ Online Safety Act 2023, c. 50 (U.K.).

⁵ Shaip. (2023, 20 сентября). 5 типов модерации контента и как масштабировать ее с помощью ИИ? <https://ru.shaip.com/blog/5-types-of-content-moderation/>

к скорости взаимодействия пользователей и необходимостью минимизации юридических и репутационных рисков⁶.

Алгоритмические методы модерации применяются для автоматического выявления неподходящего контента, эффективно фильтруя явные нарушения, такие как нецензурная лексика и «стоп-слова».

Этические и правовые требования к алгоритмической модерации включают принципы прозрачности, объяснимости и возможности обжалования решений.

Практическая реализация этих принципов находит отражение в пользовательских соглашениях крупных платформ. Например, правила «Авито» предусматривают запрет на регистрацию нового профиля взамен заблокированного и ограничение на повторное использование номера телефона после блокировки⁷. Такие положения демонстрируют эволюцию подходов к модерации: помимо анализа контента, платформы всё чаще применяют идентификационные и поведенческие признаки для предотвращения рецидивов нарушений.

Специфика премодерации на *HR-tech*-платформах

В отличие от универсальных платформ специализированные *HR-tech*-сервисы применяют *обязательную премодерацию работодателей* как основу доверительного пространства рынка труда. Так, согласно п. 5.1.4 правил *Rabota.ru* «Компания проходит регистрацию в Сервисе. Регистрация в Сервисе Компании признается завершенной после успешного прохождения модерации»⁸. Это демонстрирует проактивный подход к формированию экосистемы: проверка легитимности работодателя осуществляется до предоставления возможности размещения вакансий.

Постмодерация как механизм защиты прав соискателей

В то время как универсальные платформы фокусируются на общих запретах, *HR-tech*-сервисы разрабатывают специальные критерии защиты трудовых прав. Например, «Авито» как сервис объявлений общего назначения устанавливает детальные правила по запрету дискриминации, включая конкретные примеры недопустимых формулировок по признакам пола, возраста, национальности, здоровья: «Не указывайте в объявлении о вакансии ограничения или преимущества, в том числе по следующим признакам: <...> “Отсутствие инвалидности — важный критерий при приеме на работу” и т. п. “Без вредных привычек”, “только для некурящих” и т. п.»⁹

Постмодерация на *HR-tech*-платформах имеет профилактическую направленность. Согласно п. 7 правил *Rabota.ru* «все вакансии и резюме проходят постмодерацию (проверяются модераторами после размещения). Вакансии и резюме, не соответствующие настоящим Правилам, могут быть заблокированы. Компаниям, размещающим вакансии, не соответствующие настоящим Правилам, может быть приостановлен или прекращен доступ к Сервису»¹⁰. Это выходит

⁶ Toncheva, I. (2025, February 19). *Types of content moderation: Benefits, challenges, and use cases*. Imagga. <https://imagga.com/blog/types-of-content-moderation-benefits-challenges-and-use-cases/>

⁷ Авито. (б.д.). *Условия использования Авито*. <https://www.avito.ru/legal/rules/terms-of-use/>

⁸ Rabota.ru. (б.д.). *Правила портала Rabota.ru*. <https://www.rabota.ru/info/>

⁹ Авито. (б.д.). *Особенности размещения в категории «Работа»*. <https://www.avito.ru/legal/job/job-listings/#discrimination>

¹⁰ Rabota.ru. (б.д.). *Правила портала Rabota.ru*. <https://www.rabota.ru/info/>

за рамки стандартной контент-модерации и создает механизм ответственности за качество предложений на рынке труда.

Важно, что уровень глубины и строгости модерации варьируется даже между специализированными *HR-tech*-платформами, что отражает разные подходы к формированию правовой культуры рынка труда. Так, согласно п. 3.7 условий использования *HeadHunter* платформа «вправе отказать Заказчику в регистрации на Сайте либо заблокировать Регистрацию на Сайте, расторгнуть Договор с Заказчиком»¹¹. Это положение демонстрирует несколько ключевых особенностей *HR-tech*-модерации.

1. Глубина верификации: *HeadHunter* не ограничивается формальной проверкой документов, а проводит комплексную оценку достоверности предоставленной информации и репутации работодателя.
2. Превентивные меры против рецидивов: особое внимание уделяется выявлению попыток обхода модерации через третьих лиц, платформа проверяет «подтверждение договорных отношений с третьими лицами (если Заказчик планирует искать персонал для третьих лиц), ранее заблокированными на Сайте».

Алгоритмическая модерация: от общих правил к отраслевой специфике

Сравнительный анализ цифровых *HR*-платформ демонстрирует качественные различия в подходах к автоматизации модерации контента, прежде всего в части выявления дискриминационных условий в вакансиях. На практике становится очевидно, что эффективная модерация не может быть сведена к формальному поиску и блокированию заранее заданных «стоп-слов» (таких как указание возраста, пола или семейного положения). Подобный подход не учитывает контекст употребления формулировок и не позволяет разграничить незаконную дискриминацию и допустимые, прямо предусмотренные законом профессиональные ограничения.

Трудовая дискриминация в российском праве определяется не самим фактом упоминания конкретного признака, а отсутствием его объективной связи с выполнением трудовой функции¹². Так, требования, основанные на поле, возрасте, состоянии здоровья или семейном положении, признаются дискриминационными, если они не обусловлены характером работы и не предусмотрены Трудовым кодексом РФ или иными нормативными актами. Следовательно, автоматизированная модерация должна опираться не только на лингвистический анализ текста вакансии, но и на понимание правовой природы соответствующих ограничений, включая нормы охраны труда, требования к медицинским осмотрам, специальные запреты и допуски, установленные законом.

Примеры из практики модерации вакансий показывают, что дискриминационные условия нередко маскируются под нейтральные или оценочные формулировки и, напротив, допустимые требования могут содержать чувствительные признаки. Так, указание на разъездной характер работы или длительные командировки само по себе не является нарушением, тогда как прямое требование «только без детей» — уже очевидный пример дискриминации. Аналогично запрет на прием женщин или лиц определенного возраста может быть законным в строго ограниченных случаях, прямо закрепленных в нормативных актах (например, при выполнении тяжелых, вредных или опасных работ либо при наличии медицинских противопоказаний).

¹¹ HeadHunter. (б.д.). *Условия использования сайтов Хэдхантер*. <https://hh.ru/terms>

¹² Трудовой кодекс Российской Федерации от 30 декабря 2001 г. № 197-ФЗ. Собрание законодательства РФ, 2002 г., № 1, ст. 3 (часть I).

Формальная фильтрация таких слов без анализа правового основания приводит либо к избыточным блокировкам, либо к пропуску незаконного контента.

Таким образом, автоматизация модерации в сфере объявлений о работе требует интеграции юридической экспертизы в алгоритмические модели. Речь идет не только о соблюдении формальных требований платформ, но и о соотношении содержания вакансии с нормами Трудового кодекса РФ, практикой их применения и целями правового регулирования. В противном случае цифровая модерация рискует утратить превентивную функцию и превратиться в формальный инструмент фильтрации, неспособный эффективно предотвращать дискриминацию на рынке труда.

Отдельного внимания заслуживают случаи, когда одно и то же условие может быть как допустимым, так и незаконным в зависимости от содержания трудовой функции. Так, требование наличия водительского удостоверения является обоснованным для водителей или работников, использующих служебный автомобиль, но утрачивает правовой смысл при включении в вакансии, не связанные с управлением транспортом. Аналогичным образом различается оценка требований к речи, внешнему виду или документам: грамотная русская речь или опрятный внешний вид допустимы как профессиональные характеристики, тогда как требования родного языка, отсутствия акцента или презентабельной внешности формируют косвенную дискриминацию.

В результате в цифровой модерации формируется особый правоприменительный слой, в котором алгоритмы не просто фильтруют текст, а фактически воспроизводят логику трудового законодательства: отделяют деловые качества от социальных, физиологических и иных характеристик, не имеющих отношения к выполнению работы. Такой подход позволяет не только снижать риски ответственности за распространение дискриминационных вакансий, но и реализовывать превентивную функцию платформ — корректировать содержание объявлений еще до возникновения спора или жалобы со стороны соискателя.

Таким образом, кейсы модерации подтверждают, что борьба с трудовой дискриминацией в цифровой среде невозможна без интеграции норм Трудового кодекса РФ в архитектуру автоматизированных решений. Исключительно формальная фильтрация «стоп-слов» не учитывает ни сложность правового регулирования, ни многообразие допустимых исключений, что делает ее недостаточной для обеспечения законности и справедливости на цифровом рынке труда.

Процессуальные гарантии и обжалование решений модерации: пределы прозрачности и предотвращение обхода правил

В научной и правоприменительной дискуссии о цифровой модерации значительное внимание уделяется вопросу процессуальных гарантий пользователей и прозрачности решений, принимаемых операторами онлайн-платформ¹³. В то же время чрезмерное упрощение данной проблематики через призму права пользователя знать причины блокировки не учитывает функциональные особенности модерации как инструмента предотвращения незаконной и недобросовестной деятельности. В условиях *HR-tech*-платформ модерация выполняет не только регулятивную, но и превентивную функцию, направленную на пресечение системных нарушений, в том числе мошенничества и вовлечения соискателей в противоправную деятельность.

¹³ См.: Kominers & Shapiro (2024). Авторы демонстрируют, что непрозрачная модерационная политика может быть более эффективной для предотвращения вредоносного контента, поскольку прозрачные правила позволяют злоумышленникам адаптировать свои сообщения для обхода фильтров.

Анализ пользовательского соглашения сервиса «Авито Подработка» для исполнителей позволяет выявить модель ограниченной процессуальной прозрачности, при которой платформа сохраняет за собой право применять меры ограничения доступа без детального раскрытия мотивов принятого решения¹⁴. Так, раздел 11 соглашения предусматривает возможность блокировки профиля при нарушении условий соглашения, наличии многократных жалоб, недобросовестных действиях исполнителя, а также в целях предотвращения нарушений законодательства. При этом соглашение не обязывает компанию в каждом случае раскрывать конкретные фактические основания блокировки или методологию их выявления.

В научной литературе ограниченная прозрачность модерационных решений критикуется как фактор, подрывающий процессуальные гарантии пользователей и усиливающий асимметрию власти между платформой и участниками цифровых отношений (Pasquale, 2015). Однако в контексте борьбы с обходом правил модерации она представляется функционально оправданной. Раскрытие детальных причин блокировки, особенно в случаях размещения объявлений с незаконным контентом либо скрытых форм вовлечения в противоправную деятельность, может создавать для нарушителей обратную связь, позволяющую адаптировать поведение и обходить установленные ограничения.

В этом смысле отказ от детализированного раскрытия причин блокировки выступает не как произвольное ограничение прав пользователя, а как элемент архитектуры противодействия злоупотреблениям. Речь идет о ситуациях, например, когда пользователь регистрируется как работодатель (например, известная компания), размещает вакансии с высокой зарплатой и минимальными требованиями к сотрудникам. А в ответ на запрос платформы прислать документы о регистрации этой компании отправляет поддельные документы. Такая компания представляет опасность для доверчивых соискателей, которые могут в надежде на трудоустройство дизайнером за 400 000 руб. в месяц выполнить бесплатные тестовые задания огромного объема, оплатить доставку каких-то материалов для тестового задания и стать жертвой иной мошеннической схемы. Большинство платформ в таких случаях блокируют нарушителя¹⁵. При повторной попытке регистрации с теми же реквизитами пользователь будет снова заблокирован во избежание повторения нарушений. Если платформа сообщит пользователю, что вакансия содержит неоправданно высокую зарплату, низкие требования к сотруднику, а представленные в подтверждение наличия компании документы откровенно поддельные, то такой нарушитель сделает вывод, как в следующий раз избежать блокировки.

Отсюда следует вывод, что платформа, обладая информацией о цифровом следе пользователя, паттернах поведения, повторяемости нарушений и использовании технических средств обхода, объективно не может раскрывать все основания принятия решения без риска компрометации механизмов модерации. Данный подход согласуется с практикой использования метаданных и поведенческих факторов в качестве самостоятельных оснований для ограничительных мер, даже при формальном соблюдении пользователем текстовых требований к размещаемому контенту.

Сравнительный анализ правил сайта *Rabota.ru* подтверждает указанную логику регулирования¹⁶. Правила сервиса прямо предусматривают право администрации блокировать вакансии,

¹⁴ Авито. (б.д.). Пользовательское соглашение сервиса «Авито Подработка». <https://www.avito.ru/legal/job/gig-user-agreement/#blocking>

¹⁵ Авито. (б.д.). Почему заблокировали профиль? <https://support.avito.ru/articles/4294#3432>; HeadHunter. (б.д.). Условия использования сайтов Хэдхантер. <https://hh.ru/terms>

¹⁶ Rabota.ru. (б.д.). Правила портала Rabota.ru. <https://www.rabota.ru/info/>

резюме или доступ к сервису без объяснения причин, а также устанавливать факт нарушения на основании учетных, статистических и иных данных сервиса либо при получении определенного количества жалоб. При этом пользователь фактически соглашается с тем, что оценка достаточности и обоснованности таких данных осуществляется администрацией сервиса по собственному усмотрению. Такая конструкция, с одной стороны, ограничивает возможности пользователя по немедленному оспариванию конкретных мотивов блокировки, а с другой — существенно снижает риски воспроизводства мошеннических схем за счет обучения нарушителей на результатах модерации.

Следует подчеркнуть, что речь не идет о полном отказе от процессуальных гарантий. И «Авито Подработка», и *Rabota.ru* закрепляют возможность обращения пользователя в службу поддержки и применения претензионного порядка разрешения споров. Однако данные механизмы носят преимущественно корректирующий, а не раскрывающий характер: их цель заключается не в предоставлении пользователю исчерпывающей информации о логике модерации, а в проверке обоснованности меры в конкретном случае и исправлении возможной ошибки. Таким образом, апелляция в рамках платформы ориентирована на восстановление прав добросовестного пользователя, а не на деконструкцию внутренних алгоритмов выявления нарушений.

Итого, полная прозрачность стратегий модерации создает у нарушителей инструмент для адаптации и обхода систем модерации.

Экономическая и стратегическая логика *opaque moderation*

Это подтверждается и в зарубежной научной литературе, где концепция модерации с непрозрачными правилами (*opaque policies*) обсуждается как обоснованная стратегия платформ. В исследовании С. Д. Коминерса и Дж. М. Шапиро анализируется, как стратегия модерации, при которой модератор не раскрывает правила блокировки сигнала (контента), может быть более эффективной в ограничении распространения вредоносных сведений или дезинформации (Kominers & Shapiro, 2024). Авторы показывают, что если политика модерации прозрачна, те, кто стремится разместить вредоносный контент, могут предсказать реакцию системы и адаптировать свои сообщения так, чтобы они обходили фильтры. В случае непрозрачной модерации, когда пользователи не знают, какие именно сигналы будут заблокированы, уровень обхода систем значительно ниже.

В более новом проекте авторы развивают идеи *Robust Content Moderation*: модератор, который может удалять или фильтровать контент, должен действовать таким образом, чтобы максимально снижать вероятность вредоносного воздействия, даже если это означает закрытие части информации о причине таких действий (Kominers & Shapiro, 2024).

Анализ *algospeak* — языка обхода алгоритмов, когда пользователи сознательно изменяют словоформы и выражения, чтобы избежать автоматического удаления или понижения контента, — также подтверждает экономическую мотивацию ограниченной прозрачности модерации. Это эмпирически подтверждает, что злоумышленники адаптируются к известным правилам, внедряя обходные трюки в тексты, что снижает эффективность полностью прозрачных модерационных стратегий¹⁷.

При этом зарубежное регулирование также исходит из необходимости баланса между прозрачностью и безопасностью. Так, в *Digital Services Act* закреплена обязанность онлайн-платформ

¹⁷ Algospeak. (2025, September 21). In *Wikipedia*. <https://en.wikipedia.org/wiki/Algospeak>

представлять пользователям так называемые *Statements of Reasons* — уведомления о причинах удаления контента или ограничения доступа¹⁸. Однако данный механизм не предполагает раскрытия внутренних алгоритмов, весов критериев или технических параметров модерации.

В этом контексте положения пользовательских соглашений, допускающие отсутствие детализированного объяснения причин блокировки объявлений с незаконным или подозрительным контентом, не могут рассматриваться как произвольные или лишенные рационального основания. Напротив, они снижают возможность системного обхода ограничений. Процессуальные гарантии пользователя в такой модели смещаются с требования полного раскрытия причин блокировки на обеспечение доступных механизмов апелляции, проверки ошибок и восстановления доступа в случаях добросовестного поведения, что позволяет достичь баланса между защитой прав пользователей и устойчивостью платформенной экосистемы.

Наконец, исследования, посвященные *shadowbanning* — скрытой форме модерации, при которой пользователи не уведомляются прямо о причинах приостановки видимости их контента, показывают, что платформы сознательно создают такие непубличные подходы, чтобы не давать обратной связи, которую могли бы использовать злоумышленники для корректировки поведения (Risius & Blasiak, 2024, pp. 817–829).

Практические риски полной прозрачности

На практике раскрытие причин блокировки может дать злоумышленнику «черный ящик» с правилами, который он использует для систематического обхода модерации. В *HR-tech*-среде это может выглядеть так: если платформа указывает, что блокировка произошла из-за конкретного сочетания слов в тексте или определенной структуры объявления (например, определенные условия для соискателей, заведомо дискриминационные формулировки, призывы к незаконной деятельности и др.), то явный перечень таких критериев становится инструкцией по обходу. Пользователь может изменить номер телефона, переформулировать текст, сохранить скрытые схемы коммуникации в личных сообщениях — и таким образом обходить фильтры, не нарушая формально видимого правила.

Это особенно актуально для *HR-tech*-платформ, где целевые действия злоумышленников носят экономический характер: необходимость постоянного привлечения соискателей для незаконной деятельности, мошеннических схем или вовлечения людей в условия, противоречащие законодательству о рынке труда. В таком контексте платформа, сообщая злоумышленникам полные основания блокировки, по существу предоставляет им инструменты по обходу системы, подрывая цель модерации как защитного механизма.

Баланс между прозрачностью и безопасностью

Это не означает полного отказа от принципов прозрачности; скорее речь идет о балансе между правами пользователя и эффективностью систем защиты. Пользователь должен иметь возможность обжаловать конкретные решения и понимать, что его заблокировали, если это произошло ошибочно, но развернутые алгоритмы блокировки и внутренняя логика платформы, как правило, остаются закрытыми от публичного доступа. Такая модель согласуется с положениями *Digital Services Act* по вопросу *Statements of Reasons* — обязательных для крупных платформ объяснениях причин удаления конкретного контента. Тем не менее этот стандарт применяется в рамках строгих требований к объему и типу раскрываемой

¹⁸ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 Oct. 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), art. 17, 2022 O.J. (L 277) 1.

информации, не обязывая показывать все внутренние алгоритмы (Drolsbach & Pröllochs, 2024, pp. 939–942).

Таким образом, условия пользовательских соглашений, которые позволяют не раскрывать причин блокировки в ряде случаев, не только не лишены здравого смысла, но и могут рассматриваться как стратегическая обязанность платформы по предотвращению обхода модерации и снижению рисков злоупотреблений. В *HR-tech*-экосистемах это особенно важно, поскольку речь идет о доступе человека к экономическим возможностям и рисках вовлечения в недобросовестные схемы. Здесь высокая адаптивность злоумышленников требует умелого сочетания ограниченной прозрачности и доступных механизмов апелляции для добросовестных пользователей.

Выводы

Цифровая модерация контента на *HR-tech*-платформах выступает самостоятельным правовым механизмом, обеспечивающим качество вакансий, защиту прав соискателей и формирование безопасного цифрового пространства на рынке труда.

HR-tech-платформы реализуют модерацию как в рамках публично-правовых обязанностей (ст. 10.7 Закона № 149-ФЗ), так и через добровольные внутренние стандарты, создавая сочетание регулирования и саморегулирования, что позволяет эффективно управлять рисками недобросовестной деятельности пользователей.

Модерация на *HR-tech*-платформах выполняет функцию первичной юридической квалификации фактов в цифровой среде: блокировка контента и ограничение доступа пользователей формируют правовые последствия, опережающие вмешательство государственных органов, и служат инструментом превентивного регулирования.

Эффективность модерации обеспечивается балансом между прозрачностью и безопасностью: ограниченное раскрытие причин блокировки снижает возможности обхода системы злоумышленниками, причем пользователи сохраняют процессуальные гарантии через апелляцию и корректирующие процедуры.

Специализированные алгоритмические и процессуальные подходы к модерации на *HR-tech*-платформах интегрируют отраслевую специфику трудовых отношений и нормы трудового законодательства, что обеспечивает дифференцированную защиту участников цифрового рынка труда и повышает доверие к платформе.

В эссе систематизировано и концептуализировано правовое значение цифровой модерации на *HR-tech*-платформах как механизма первичной юридической квалификации и опережающего регулирования. Подчеркнута роль ограниченной прозрачности модерационных решений как стратегической меры противодействия злоупотреблениям.

Список литературы / References

1. Altukhov, A. V., & Kashkin, S. YU. (2021). Pravovaya priroda tsifrovyykh platform v rossiyskoy i zarubezhnoy doctrine [Legal nature of digital platforms in Russian and foreign doctrine]. *Actual Problems of Russian Law*, (7), 86–94.
2. Casas-Cortés, M., Cañedo-Rodríguez, M., & Diz, C. (2023). Platform capitalism. In M. Aldenderfer (Ed.), *Oxford research encyclopedia of anthropology*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190854584.013.597>

3. Drolsbach, C., & Pröllochs, N. (2023). *Content moderation on social media in the EU: Insights from the DSA Transparency Database* (arXiv:2312.04431). arXiv. <https://doi.org/10.48550/arXiv.2312.04431>
4. Khabrieva, T. Ya., & Chernogor, N. N. (2018). Pravo v usloviyakh tsifrovoy real'nosti [Law in the conditions of digital reality]. *Journal of Russian Law*, (1), 85–102. https://doi.org/10.12737/art_2018_1_7
5. Kominers, S. D., & Shapiro, J. M. (2024). *Content moderation with opaque policies* (NBER Working Paper No. 32156). National Bureau of Economic Research. <https://doi.org/10.3386/w32156>
6. Krasavchikov, O. A. (1958). *Yuridicheskiye fakty v sovetskom grazhdanskom prave*. Gosyurizdat.
7. Pasquale, F. (2015). *The black box society, the secret algorithms that control money and information*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674736061>
8. Risius, M., & Blasiak, K. M. (2024). Shadowbanning. *Business & Information Systems Engineering*, (66), 817–829. <https://doi.org/10.1007/s12599-024-00905-3>

Сведения об авторе:

Шабронова С. Е. — старший юрист, ООО «Хэдхантер», Москва, Россия.

s.shabronova@yandex.ru

ORCID: <https://orcid.org/0009-0006-4502-7903>

Information about the author:

Svetlana E. Shabronova — Senior Lawyer, HeadHunter LLC, Moscow, Russia.

s.shabronova@yandex.ru

ORCID: <https://orcid.org/0009-0006-4502-7903>