Право первой публикации данной статьи предоставлено ее автором журналу открытого доступа «Legal Issues in the Digital Age». ISSN электронной версии: 2713-2749

Дискриминационный каскад в системах искусственного интеллекта публичного права: правовые механизмы противодействия

УДК 342.7
DOI: [будет присвоен редакцией]

Фролов Александр Александрович
ORCID iD: 0009-0008-7131-1785
Аспирант, Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева, Саранск, Россия
Руководитель юридической службы АО «Орбита» г. Саранск
e-mail: afrolov77@yandex.ru

Аннотация

Исследование посвящено разработке правовой модели противодействия дискриминационному каскаду в системах искусственного интеллекта на основе жизненно-цикловой рамки источников вреда. Концепция дискриминационного каскада описывает механизм, при котором отдельные источники предвзятости и ошибок на разных стадиях жизненного цикла ИИ-систем наслаиваются друг на друга и взаимно усиливают дискриминационные эффекты в конечных решениях. На основе компаративно-правового метода, критического анализа нормативных актов и социотехнического анализа жизненного цикла ИИ выявлены семь уровней каскадного накопления дискриминационных эффектов. Сравнительный анализ российских и зарубежных подходов показывает, что западная доктрина предлагает формализованную жизненно-цикловую рамку и

операционализацию каскадной модели, тогда как российские исследователи концентрируются на правовых рисках и конституционном принципе равенства. Критический анализ выявил системные пробелы российского регулирования: отсутствие требований к репрезентативности данных, стандартов аудита на дискриминацию по всем этапам жизненного цикла, механизмов мониторинга и процедур прерывания обратных связей. Особое внимание уделено рискам дискриминационного каскада в контексте цифровизации контрольно-надзорной деятельности. Сформулирована многоуровневая модель правового вмешательства, основанная на принципах pipeline-aware governance. Результаты исследования могут быть использованы при разработке комплексного федерального закона об искусственном интеллекте в Российской Федерации.

Ключевые слова: дискриминационный каскад, искусственный интеллект, жизненный цикл ИИ, алгоритмическая дискриминация, pipeline-aware fairness, контрольно-надзорная деятельность, источники вреда, многоуровневое вмешательство, EU AI Act

Discriminatory cascade in public law artificial intelligence systems: legal counteraction mechanisms

Abstract

The study develops a legal model for countering discriminatory cascades in AI systems based on a life-cycle framework of sources of harm. The concept describes a mechanism whereby individual bias and error sources at different AI lifecycle stages layer upon and mutually reinforce discriminatory effects in final decisions. Using comparative legal method, critical normative analysis, and sociotechnical AI lifecycle analysis, seven cascading levels of discriminatory effect accumulation were identified. Western doctrine offers a formalized life-cycle framework and cascade model operationalization, while Russian researchers focus on legal risks and constitutional equality. Critical analysis revealed systemic Russian regulatory gaps: lack of data representativeness requirements, discrimination auditing standards across lifecycle stages, post-market monitoring, and feedback loop

interruption procedures. Special attention addresses discriminatory cascade risks in digitalization of control and supervisory activities. A multi-level legal intervention model based on pipeline-aware governance principles is formulated. Results support comprehensive federal AI legislation development in the Russian Federation.

Keywords: discriminatory cascade, artificial intelligence, AI life cycle, algorithmic discrimination, pipeline-aware fairness, control and supervisory activities, sources of harm, multi-level intervention, EU AI Act

ВВЕДЕНИЕ

Интеграция систем искусственного интеллекта в публичное управление приобретает масштабный характер. В 2025 году Министерство экономического развития Российской Федерации инициировало внедрение ИИ-технологий в контрольно-надзорную деятельность, ориентируясь на переход к риск-ориентированному подходу. Параллельно нарастает осознание дискриминационного потенциала алгоритмов, способных усиливать исторические предубеждения через механизм автоматической дискриминации [Талапина, 2025: 55]. Принятие Регламента (ЕС) 2024/1689 об искусственном интеллекте в июне 2024 года установило международный стандарт риск-ориентированного регулирования, признающего многоуровневую природу алгоритмической дискриминации[1]. Рамочная конвенция Совета Европы об искусственном интеллекте и правах человека, демократии и верховенстве права (CETS No. 225), открытая для подписания в сентябре 2024 года, дополнительно подчёркивает значимость данной проблематики для международного правового пространства[2].

Предметом настоящего исследования выступает концепция дискриминационного каскада как модель мультифакторного и фазового

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) // Official Journal of the European Union. 2024. L 206. 12 July.

[2] Council of Europe. Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225) : opened for signature 5 September 2024.

возникновения дискриминации в системах искусственного интеллекта. В отличие от упрощённой схемы «смещённые данные - смещённый результат», каскадная модель акцентирует, что причинно значимые источники вреда распределены по всему жизненному циклу ИИ-систем: от постановки задачи до формирования обратных связей. Каждый этап не просто передаёт смещение дальше, но способен его трансформировать, структурировать и умножать. Данная модель представляет собой не только теоретическую конструкцию, но и аналитический инструмент, позволяющий проектировать адресные правовые интервенции.

Целью работы является разработка правовой модели противодействия дискриминационному каскаду в ИИ-системах на основе жизненно-цикловой рамки, предложенной Suresh и Guttag [Suresh, Guttag, 2021: 1-15]. Для достижения цели решаются следующие задачи:

- концептуализировать дискриминационный каскад через призму жизненного цикла ИИ;
- идентифицировать уровни и механизмы каскадного накопления дискриминационных эффектов;
- провести сравнительный анализ подходов российских и зарубежных исследователей;
- выявить пробелы российского регулирования; сформулировать правовые механизмы многоуровневого вмешательства.

Методологическую основу составляет комбинация компаративно-правового метода, критического анализа нормативных актов (Указ Президента РФ от 10 октября 2019 г. No 490, Федеральный закон от 24 апреля 2020 г. No 123-ФЗ, Федеральный закон от 8 августа 2024 г. No 233-ФЗ, Регламент (ЕС) 2024/1689, Регламент (ЕС) 2016/679), социотехнического анализа жизненного цикла ИИ и концептуального моделирования[3].

[3] Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» // СЗ РФ. 2019. № 41. Ст. 5700. ; Федеральный закон от 8 августа 2024 г. № 233-ФЗ «О внесении изменений в Федеральный закон “О персональных данных”» // Собрание законодательства Российской Федерации. 2024. № 33. Ст. 4593. ; Федеральный закон от 24.04.2020 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного

Теоретическую базу формируют работы в области «fairness in machine learning», социотехнического анализа алгоритмических систем и правового регулирования ИИ.

1. Концептуальная рамка дискриминационного каскада

Традиционное понимание алгоритмической предвзятости часто сводится к линейной схеме «bias in, bias out», в рамках которой проблема дискриминации локализуется исключительно на этапе данных. Такая редукция не позволяет уловить системную природу алгоритмической дискриминации и объясняет ограниченность локальных технических интервенций. Харитонова, Савина и Паньини справедливо отмечают, что алгоритмическая предвзятость выявляет скрытые структурные неравенства общества [Харитонова, Савина, Паньини, 2021: 492]. Однако системы искусственного интеллекта не только выявляют, но и структурируют и усиливают эти неравенства через каскадный механизм, при котором локальные решения в данных, дизайне и внедрении создают совокупный системный эффект, превышающий сумму отдельных смещений.

Suresh и Guttag предложили жизненно-цикловую рамку (framework), систематизирующую семь типов источников вреда на разных этапах ML-пайплайна [Suresh, Guttag, 2021: 1-15]. Каждый тип соответствует определённому этапу жизненного цикла:

- historical bias отражает предвзятость, встроенную в социальный контекст и запечатлённую в данных;
- representation bias это недопредставленность определённых групп в обучающей выборке;
- measurement bias отражает искажения при операционализации целевых переменных и признаков;

интеллекта в субъекте Российской Федерации — городе федерального значения Москве» // СЗ РФ. 2020. № 17. Ст. 2701. ; Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) // Official Journal of the European Union. 2016. L 119. P. 1–88.

-aggregation bias отражает потерю групповой специфики при использовании единой модели для разнородных популяций;

- learning bias отражает усиление смещений в процессе оптимизации;

- evaluation bias это несбалансированность тестовых данных и метрик оценки;

- deployment bias это несоответствие контекста практического применения условиям разработки и обучения.

Эта типология принципиально важна тем, что демонстрирует распределённость источников вреда по всей цепочке создания и использования ИИ-систем.

Автором настоящего исследования выдвигается концепция «дискриминационного каскада» как механизм анализа и минимизации алгоритмической дискриминации. Дискриминационный каскад в авторском понимании - концептуальная модель, раскрывающая многоуровневый механизм возникновения, накопления и усиления дискриминационных эффектов на различных этапах жизненного цикла ИИ-системы. Модель состоит из четырёх последовательных этапов:

Этап 1 - подготовка и отбор данных («входящая дискриминация»): Обучающие данные отражают исторические закономерности, включая историческую дискриминацию против определённых групп. Например, архивы судебных решений могут содержать неравное распределение приговоров по полу, национальности или социальному статусу. ИИ-система, обучаясь на этих данных, неизбежно усваивает эти предубеждения.

Этап 2 - обучение модели и калибровка параметров («усиление предубеждений»). Алгоритм не просто воспроизводит предубеждения входящих данных, а активно их усиливает. Это происходит потому, что алгоритм оптимизируется на минимизацию ошибок предсказания в целом, но может при этом непропорционально увеличивать ошибки для меньшинства. Например, если в обучающих данных 90% решений для мужчин и 10% для женщин, алгоритм может «специализироваться» на правильном предсказании мужских случаев в ущерб женским.

Этап 3 - развёртывание системы и применение к реальным субъектам («распространение дискриминации»):

Обученная и усиленная модель применяется к новым людям в реальных судебных и административных процессах. На этом этапе дискриминационные ошибки становятся судебными решениями или административными актами, имеющими реальные последствия для граждан.

Этап 4 - мониторинг и обратная связь («закрепление дискриминации»): Ошибочные решения и их последствия часто становятся частью нового обучающего набора при переобучении модели (retrain). Это создаёт обратную связь: дискриминационные ошибки прошлого фиксируются в системе как стандарты будущего, ещё больше закрепляя дискриминационные паттерны.

Юридическая полезность предлагаемой концепции заключается в том, что концепция дискриминационного каскада решает четыре практических задачи:

1. Задача определения ответственного (Источник ответственности). То есть на каком этапе возникла проблема дискриминации, и кто за неё ответственен:

- если дискриминация выявлена на Этапе 1, ответственен разработчик или аналитик данных, который не провел проверку исходных данных на предмет дисбаланса;
- если на Этапе 2 - ответственен разработчик модели, который не использовал методы для смягчения дискриминации (fairness-aware algorithms);
- если на Этапе 3 - ответственен оператор системы (суд или надзорное ведомство), которое не провело независимый аудит перед внедрением;
- если на Этапе 4 - ответственность может быть разделена, но именно оператор должен был вывести систему из эксплуатации при обнаружении проблемы.

2. Установление стандартов аудита: каждый этап должен быть проверен с использованием специальных метрик справедливости (fairness metrics).

3. Судебные последствия. При выявлении дискриминационного каскада могут наступить следующие судебные и административные последствия:

- отмена судебных решений, принятых с использованием дискриминационной системы;
- обязание органа юстиции провести корректирующие мероприятия и переучить систему;
- обязание перепроверить дела, рассмотренные с использованием скомпрометированной ИИ-системы;
- возмещение убытков гражданам, пострадавшим от дискриминационных решений.

4. Применение при разработке ИИ-систем. На этапе разработки концепция дискриминационного каскада позволяет:

- Предвидеть потенциальные точки уязвимости;
- Внедрять профилактические меры на каждом этапе (сбалансированные данные, fairness-aware algorithms, аудит перед развёртыванием, постоянный мониторинг);
- Документировать все проведённые проверки для демонстрации добросовестности разработчика.

Рекомендуемые автором области применения разработанной концепции:

- Системы оценки риска рецидива в уголовном судопроизводстве (в контексте определения меры пресечения, условно-досрочного освобождения);
- Системы предиктивной аналитики для распределения судебных дел;
- Системы автоматизированного анализа судебных архивов;
- Системы, используемые органами внутренних дел для оценки рисков и принятия решений о мерах пресечения.

Пример реализации концепции в судебной практике:

В 2016 году в США система COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) для оценки риска рецидива стала предметом общественного разбирательства. Исследование журналистов

ProPublica выявило, что система демонстрирует расовую предвзятость: уровень ложноположительных ошибок (false positive rate) для афроамериканских подсудимых составляет 45%, а для белых - 23%[4]. Это означает, что 45% афроамериканцев, которые не совершат повторное преступление, были ошибочно классифицированы как высокорисковые; аналогичный показатель для белых составляет 23%.

Анализ через призму дискриминационного каскада показывает:

- Этап 1: исходные данные о преступлениях содержали исторический дисбаланс в аресте и обвинении афроамериканцев (из-за systemic racism в правоохранительной системе);
- Этап 2: алгоритм усилил этот дисбаланс, найдя корреляции между расовыми признаками и предсказанием рецидива;
- Этап 3: система применялась в реальных судебных процессах, рекомендуя строже относиться к афроамериканским подсудимым;
- Этап 4: повышенные уровни рецидива среди афроамериканцев (следствие более строгих мер) вновь вводились в обучающий набор, закрепляя дискриминацию.

Это дело служит предостережением для Российской Федерации: при внедрении ИИ-систем в уголовное судопроизводство необходимо заранее провести проверку на предмет социальной, национальной и половой дискриминации, используя концепцию дискриминационного каскада.

Концепция дискриминационного каскада, развиваемая в настоящей работе, подчёркивает накопительный и усиливающий характер дискриминационных эффектов. На социально-структурном уровне исторически сложившиеся неравенства формируют исходное распределение событий, которое «запечатывается» в данных. Barocas, Hardt и Narayanan подчёркивают, что данные не нейтральны - они отражают институциональные практики, включая практики неравного обращения [Barocas, Hardt, Narayanan, 2023]. На уровне сбора данных недопредставленность групп, выбор удобных

[4] Angwin J., Larson J., Mattu S., Kirchner L. Machine Bias [Электронный ресурс] // ProPublica. 23 May 2016. URL: https://www.propublica.org/article/machine-bias (дата обращения: 13.11.2025).

источников и историческая селекция формируют фундамент искажений. Маркировка и категоризация данных неизбежно отражают доминирующие предубеждения разработчиков и заказчиков [Харитонова, Савина, Паньини, 2021: 497].

На этапе предобработки данных выбор признаков может косвенно кодировать защищаемые основания через прокси-переменные: почтовый индекс используется вместо этничности, наименование образовательного учреждения вместо социально-экономического статуса. Самбасиван и соавторы продемонстрировали, что выбор прокси-переменных и их сочетаний существенно различается в зависимости от социокультурного контекста, что затрудняет перенос западных fairness-решений в иные юрисдикции [Sambasivan et al., 2021: 315-328]. Структурная дискриминация воспроизводится и через выбор архитектуры модели и целевой функции: оптимизация под общую точность (accuracy) игнорирует распределение ошибок между группами, поощряя стратегию «жертвуем меньшинством ради общего показателя» [Харитонова, Савина, Паньини, 2021: 500]. Heidari и соавторы формализовали эту проблему через индексы неравенства, продемонстрировав, что различные метрики справедливости отражают различные нормативные позиции относительно приемлемого распределения ошибок [Heidari et al., 2018: 2239-2248].

На этапе оценки модели выбор метрик может скрывать групповые провалы: высокая общая точность способна маскировать катастрофическую частоту ошибок для малочисленной группы [Suresh, Guttag, 2021: 1-15]. Исследование кумулятивного эффекта множественных fairness-интервенций, установило, что последовательное применение корректирующих мер на разных этапах пайплайна может приводить к непредсказуемым и неаддитивным результатам [Ghai B., Mishra M., Mueller K., 2022: 3875-3885]. Это подтверждает каскадную природу проблемы: интервенция на одном уровне влияет на эффективность интервенций на других уровнях.

На этапе внедрения статистическая предвзятость трансформируется в юридически релевантную дискриминацию. Талапина указывает, что дискриминационный потенциал реализуется именно при применении алгоритмов в принятии решений, затрагивающих права граждан [Талапина, 2025: 56–57]. Косвенная дискриминация через формально нейтральные критерии становится основным каналом каскадного вреда. Holstein и соавторы, проведя серию интервью с практиками индустрии, установили, что разрыв между средой разработки и средой внедрения является одним из главных факторов дискриминационных последствий ИИ-систем [Holstein et al., 2019: 1-16].

На уровне обратной связи алгоритмические решения изменяют поведение людей и институтов, формируя новые смещённые данные и создавая самоусиливающиеся циклы. Классический пример - предиктивная полицейская аналитика, где усиленное патрулирование определённых районов порождает больше арестов, которые «подтверждают» высокий риск и замыкают петлю обратной связи. Selbst и Barocas показали, что такие петли обратной связи представляют особую сложность для антидискриминационного права, поскольку каждое конкретное решение в цикле может представляться обоснованным, тогда как дискриминационным является системный паттерн [Selbst, Barocas, 2016: 671-732].

Логика каскада принципиально отличается от линейной модели «bias in / bias out» мультифакторностью и фазовостью. Каждый этап не просто передаёт, но трансформирует, структурирует и умножает дискриминацию. Рассмотрим, как это может выглядеть на примере риск-ориентированного надзора в сфере охраны окружающей среды и санитарных требований для предприятий[5]:

[5] См., напр.: Systemic Enforcement Bias → Term // Pollution & Sustainability Directory. URL: https://pollution.sustainability-directory.com/term/systemic-enforcement-bias/; Kuehn R. R. Bias in Environmental Agency Decision Making // Environmental Law. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2585173; https://law.lclark.edu/live/files/21023-45-4kuehnpdf; Algorithmic Discrimination: Examining Its Types and Regulatory Measures // Frontiers. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC11148221/; When Algorithms Import Private Bias into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions // Stanford RegLab. URL: https://reglab.stanford.edu/publications/when-algorithms-import-private-bias-into-public-enforcement/; Campaign-style Enforcement and Corporate Environmental Compliance // Frontiers in Public Health. URL: https://www.frontiersin.org/journals/public-health/articles/10.3389/

1. Исторический и репрезентативный bias. Исторические данные о нарушениях формируются в условиях неравномерного применения надзора: неангажированные и экономически слабые компании чаще становятся объектом проверок и санкций, крупные игроки - чаще избегают жёсткого реагирования.

В результате в реестрах нарушений и делах об административных правонарушениях «перепредставлены» малые и региональные предприятия, а крупные системные загрязнители или нарушители градостроительных норм недоотражены.

2. Измерительный bias при сборе данных. В качестве «сигналов риска» используются жалобы граждан, медиа-упоминания, открытые данные и другие сетевые источники, но каналы жалоб и освещения более доступны и активны в районах с ресурсными и организованными сообществами, тогда как бедные и маргинализированные районы жалуются меньше. Это приводит к тому, что интенсивность сигналов о нарушениях фиксируется как более высокая для уже проблемных территорий и небольших объектов, в то время как системные, но политически защищённые источники риска оказываются «тихими» в данных.

3. Агрегационный bias в моделях риска. На основании исторических данных регулятор разрабатывает единую риск-модель для всех предприятий отрасли (например, по параметрам размера, типа производства, географии), не разделяя группы по институциональному контексту, доступу к правовой защите, политическому весу. Модель воспринимает «частоту выявленных нарушений» как универсальный индикатор риска, агрегируя гетерогенные ситуации и закрепляя повышенный риск-класс за теми категориями предприятий, которые исторически чаще попадали в поле надзора, а не за теми, кто объективно наносит больший ущерб.

4. Learning bias при настройке системы. Оптимизация риск-модели проводится по общей эффективности: максимизируется «попадание»

---

fpubh.2025.1688719/full.

проверок в уже ожидаемые нарушения, с опорой на прошлые данные о результативности инспекций (сколько нарушений выявлено на 1 проверку). Это поощряет концентрацию ресурса там, где нарушения «гарантированно» выявляются (малый бизнес, слабозащищённые субъекты), и не стимулирует исследование недоисследованных сегментов (крупные загрязнители, инфраструктурные проекты, политически чувствительные объекты).

5. Evaluation bias в мониторинге эффективности. Эффективность надзорной деятельности оценивается агрегированными показателями: количество выявленных нарушений, суммарный объём штрафов, доля проверок с установленными правонарушениями и др., без разбиения по типам субъектов и их социальному положению. На несбалансированной выборке (почти все проверки - у слабых игроков) система показывает «успех» (высокий процент результативных проверок), а систематический «недоохват» более влиятельных нарушителей остаётся статистически невидимым.

6. Deployment bias в практическом применении. Риск-модель используется при планировании проверок и кампаний: субъекты с высоким баллом риска включаются в ежегодные планы, становятся объектом «кампаний усиленного контроля», при этом решения инспекторов и руководства де-факто следуют рекомендациям системы. У предприятий из уязвимых категорий возрастает частота проверок, объём требований и санкций, тогда как крупные игроки, классифицированные как «низкий риск», фактически получают режим регуляторной благосклонности и меньшую вероятность жёстких мер.

7. Кумулятивный эффект для уязвимых субъектов. Повышенная нагрузка надзора (штрафы, предписания, простои) для средних, малых и микропредприятий снижает их устойчивость и конкурентоспособность, что закрепляет их положение «проблемных» субъектов и повышает вероятность последующих нарушений (например, из-за экономии на экологических или санитарных мерах). Новые эпизоды нарушений снова попадают в данные и подтверждают риск-оценку, усиливая асимметрию надзора и закрепляя системный enforcement bias: формально рациональные решения на каждом

этапе (использование жалоб, метрик эффективности, риск-ориентированности) в совокупности создают несправедливо более жёсткий надзор над уязвимыми субъектами и мягкий режим для структурно сильных.

2. Сравнительный анализ подходов к каскадной модели

Зарубежная доктрина выработала развитый инструментарий для концептуализации многоуровневой природы алгоритмической дискриминации. В центре этого инструментария - явная жизненно-цикловая рамка, структурирующая анализ источников вреда по этапам.

В правовом измерении Wachter, Mittelstadt и Floridi исследовали, как алгоритмическая дискриминация соотносится с нормами GDPR и недискриминационным правом Европейского союза, обосновывая, что правовая оценка должна охватывать всю цепочку создания и применения ИИ-систем [Wachter, Mittelstadt, Floridi, 2017: 76-99]. Grozdanovski L. демонстрирует, что правовые инструменты ЕС - Регламент о защите персональных данных (GDPR)[6], Регламент об искусственном интеллекте (AI Act)[7] и нормы о недискриминации - в совокупности создают правовую экосистему, позволяющую оспаривать алгоритмическую дискриминацию, но каждый из инструментов в отдельности покрывает лишь часть каскада [Grozdanovski L., 2025: 5039-5062].

Российская доктрина рассматривает алгоритмическую дискриминацию преимущественно через призму конституционного принципа равенства и правовых рисков цифровизации. Талапина последовательно разрабатывает концепцию рисков дискриминации при обработке данных с помощью ИИ, связывая их с алгоритмизацией государственного управления [Талапина, 2022: 4–27; 2025: 55-58]. Её работы акцентируют, что автоматизация

[6] Регламент Европейского Парламента и Совета Европейского Союза 2016/679 от 27 апреля 2016 г. «О защите физических лиц при обработке персональных данных и о свободном обращении таких данных, а также об отмене Директивы 95/46/ЕС (Общий регламент о защите персональных данных / General Data Protection Regulation / GDPR)» // Официальный журнал Европейского Союза. 2016. № L 119. С. 1–88.

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) // Official Journal of the European Union. 2024. L 206. 12 July.

государственных функций с использованием ИИ создаёт риски нарушения конституционного принципа равенства, при этом российское законодательство не содержит адекватных инструментов предотвращения таких нарушений. Талапина вводит понятие «дискриминационного потенциала алгоритмов», описывающее латентную способность алгоритмических систем воспроизводить и усиливать неравенства, что методологически сближает её подход с концепцией каскада [Талапина, 2025: 55–58].

Харитонова, Савина и Паньини подчёркивают, что предвзятость алгоритмов - отражение структурных неравенств общества, требующее комплексного правового ответа на стыке этики и права [Харитонова, Савина, Паньини, 2021: 488-515]. Смирнова анализирует этические аспекты алгоритмической предвзятости, подчёркивая значение культурного контекста для понимания и оценки дискриминационных эффектов [Смирнова, 2023: 118-126]. Эти понятия концептуально близки к каскадной модели, хотя авторы не использует явно жизненно-цикловую рамку. Глушкова и Летунов рассматривают цифровые права нового поколения, формирующиеся в ответ на вызовы цифровизации, включая право на недискриминацию в алгоритмических решениях [Глушкова, Летунов, 2020: 16-28].

Ключевые различия между подходами проявляются по нескольким осям. По степени формализации модели: зарубежные работы предлагают структурированные модели по этапам ML-цикла, позволяющие проектировать интервенции «по слоям»; в российской литературе концепция каскада либо вообще не учитывается, либо каскад выступает как концептуальная связка блоков риска без детальной «карты» по всем стадиям, а систематизация строится вокруг правовых категорий. По фокусу анализа зарубежные исследователи сильнее акцентируют социотехнический аспект взаимодействия технических и институциональных факторов. Российские авторы концентрируются на юридико-доктринальных последствиях и конституционных основаниях. По привязке к регуляторным инструментам

зарубежные подходы непосредственно связаны с конкретными регуляторными механизмами (GDPR, AI Act, антидискриминационное право), предлагая многоуровневые механизмы управления рисками. Российские работы чаще ограничиваются нормативно-правовым анализом существующего законодательства и формулированием общих рекомендаций.

Вместе с тем точки соприкосновения значительны: обе исследовательские традиции признают многоуровневую природу алгоритмической дискриминации, недостаточность локальных технических решений и необходимость комплексного правового регулирования. Ключевое различие состоит в степени операционализации: зарубежный подход предоставляет «технический скелет» каскада с формализованными метриками и стадиями, а российский - «правовой контекст» конституционных принципов, в который этот скелет может быть встроен [Талапина, 2022: 4-27; 2025: 55-58]. Интеграция обоих подходов представляется наиболее продуктивной стратегией для разработки эффективного правового регулирования.

3. Правовое регулирование ИИ в контексте противодействия дискриминационному каскаду

Регламент (ЕС) 2024/1689 (EU AI Act) реализует риск-ориентированный подход, классифицируя ИИ-системы по четырём уровням риска:

- запрещённые практики (социальный скоринг, манипулятивные технологии);
- высокорисковые системы (правоохранение, найм, образование, кредитование, миграционный контроль);
- системы с ограниченным риском (чат-боты, генеративный ИИ);
- системы с минимальным риском[8].

Требования к высокорисковым системам можно интерпретировать как попытку многоуровневого вмешательства в каскад:

- обязательная оценка качества данных и мер по минимизации смещений;

[8] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) // Official Journal of the European Union. 2024. L 206. 12 July.

- документирование процессов разработки и тестирования; обеспечение человеческого надзора; организация постмаркетного мониторинга [Grozdanovski L., 2025: 5039-5062].

Однако критический анализ обнаруживает существенные ограничения AI Act с точки зрения противодействия каскаду. Во-первых, Регламент недостаточно фокусируется на так называемых петлях обратной связи (feedback loops) - самоусиливающихся циклах обратной связи, которые составляют завершающий и наиболее опасный уровень каскада. Во-вторых, преобладает акцент на «ex ante» оценках без детализации требований к непрерывному мониторингу в процессе эксплуатации. В-третьих, существует риск формального соблюдения (compliance theater), когда выполнение процедурных требований не обеспечивает реального противодействия дискриминации. В работе Grozdanovski L. показывается, что полноценная защита от алгоритмической дискриминации требует совместного применения AI Act, GDPR и норм о недискриминации [Grozdanovski L., 2025: 5039-5062].

Статья 22 GDPR, запрещающая полностью автоматизированные решения с правовыми последствиями для субъекта данных, теоретически способна прервать каскад на уровне внедрения, требуя человеческого вмешательства в принятие решений[9]. Однако на практике эта норма часто обходится через формальное включение человека в процесс без обеспечения его реальной способности оценить и скорректировать алгоритмическое решение [Wachter, Mittelstadt, Floridi, 2017: 76-99]. Традиционное антидискриминационное право ЕС сфокусировано на точечном вмешательстве - оспаривании конкретного дискриминационного решения, тогда как каскад требует превентивного, структурного регулирования по всем уровням жизненного цикла.

[9] Регламент Европейского Парламента и Совета Европейского Союза 2016/679 от 27 апреля 2016 г. «О защите физических лиц при обработке персональных данных и о свободном обращении таких данных, а также об отмене Директивы 95/46/ЕС (Общий регламент о защите персональных данных / General Data Protection Regulation / GDPR)» // Официальный журнал Европейского Союза. 2016. № L 119. С. 1–88.

Соединённые Штаты применяют преимущественно сектор-специфичный подход, задействуя существующее антидискриминационное законодательство для оценки алгоритмических систем. Доктрина «disparate impact» позволяет оспаривать решения с непропорциональным воздействием на защищаемые группы. Selbst и Barocas детально исследовали применение этой доктрины к алгоритмическому принятию решений, выявив концептуальные трудности: необходимость определения релевантной «практики», каузальной связи и обоснования деловой необходимости в контексте непрозрачных алгоритмических систем [Selbst, Barocas, 2016: 671–732]. Отсутствие единого федерального регулирования ИИ в США создаёт существенные барьеры для так называемого «pipeline-aware governance»: нет систематических требований к аудиту по всем уровням каскада, а регулирование фрагментировано по секторам и штатам.

Российское регулирование искусственного интеллекта характеризуется значительной фрагментарностью. Указ Президента РФ от 10 октября 2019 г. No 490 «О развитии искусственного интеллекта в Российской Федерации» устанавливает стратегические ориентиры развития ИИ, однако не содержит операциональных норм по предотвращению дискриминации[10]. Федеральный закон от 24 апреля 2020 г. No 123-ФЗ обеспечивает точечное регулирование экспериментального режима в Москве[11]. Федеральный закон от 8 августа 2024 г. No 233-ФЗ сфокусирован на режиме персональных данных в контексте технологий ИИ[12]. ГОСТ Р 59277-2020 предлагает техническую классификацию систем ИИ без каких-либо критериев справедливости и антидискриминационных требований[13].

---

[10] Указ Президента РФ от 10 октября 2019 г. No 490 «О развитии искусственного интеллекта в Российской Федерации» // СЗ РФ. 2019. No 41. Ст. 5700.

[11] Федеральный закон от 24 апреля 2020 г. No 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве» // СЗ РФ. 2020. No 17. Ст. 2701.

[12] Федеральный закон от 8 августа 2024 г. No 233-ФЗ «О внесении изменений в Федеральный закон «О персональных данных»» // СЗ РФ. 2024. No 33. Ст. 4593.

[13] ГОСТ Р 59277-2020. Системы искусственного интеллекта. Классификация систем искусственного интеллекта : национальный стандарт Российской Федерации ; дата введения 2021-03-01. — Москва : Стандартинформ, 2020. — IV, 12 с. — (Национальный стандарт

Талапина указывает на системную неготовность российского законодательства к регулированию дискриминационного потенциала алгоритмов [Талапина, 2025: 57–58]. Пробелы проявляются на каждом уровне каскада:

- отсутствуют требования к «data governance» и репрезентативности обучающих данных;
- не установлены стандарты аудита ИИ-систем на дискриминацию по всем этапам жизненного цикла;
- не определён правовой статус косвенной дискриминации, осуществляемой через прокси-переменные;
- не урегулировано распределение ответственности между разработчиком, оператором и пользователем ИИ-системы;
- отсутствуют механизмы постмаркетного мониторинга;
- не предусмотрены процедуры выявления и прерывания обратных связей (feedback loops).

Цифровизация контрольно-надзорной деятельности (КНД) представляет особый интерес с точки зрения рисков каскада. В ноябре 2025 года Минэкономразвития Российской Федерации инициировало внедрение ИИ-технологий в КНД с целью перехода к риск-ориентированному подходу на основе алгоритмического профилирования поднадзорных субъектов. В связи с чем возникают риски деиндивидуализации и категоризации субъектов в «группы риска» на основании формальных признаков. Риск-профилирование на основе исторических данных о проверках способно закреплять систематические смещения: если определённые категории субъектов исторически проверялись чаще, данные о них будут богаче, а алгоритм воспримет это как «объективное» основание для дальнейших проверок. Использование прокси-переменных - таких как размер бизнеса, регион регистрации, отраслевая принадлежность - может косвенно дискриминировать отдельные группы предпринимателей.

Российской Федерации). — ISBN 978-5-4426-0094-3.

Механизм обратной связи (feedback loop) в КНД особенно опасен. Частые проверки, обусловленные высокой алгоритмической оценкой риска, выявляют больше нарушений (в том числе за счёт интенсивности контроля), которые «подтверждают» высокий риск в глазах системы. Система переобучается на новых данных, повышая риск-оценку данной категории субъектов. Цикл замыкается, превращая административные смещения в «объективированную» модель рисков [Талапина, 2025: 56]. При этом правовое регулирование ИИ в КНД фактически отсутствует: не установлены требования к аудиту алгоритмических систем, используемых для риск-профилирования; не определены правовые основания и ограничения алгоритмического профилирования; не предусмотрены процедурные гарантии для поднадзорных субъектов (право на информацию о применении ИИ, право на объяснение, право на оспаривание); не установлены стандарты прозрачности алгоритмов, используемых в КНД.

4. Правовое механизмы противодействия: многоуровневая модель

Из каскадного понимания алгоритмической дискриминации вытекает необходимость правового регулирования, охватывающего все стадии жизненного цикла ИИ-систем. Харитонова, Савина и Паньини подчёркивают необходимость комплексного подхода, предусматривающего правовое воздействие на всех стадиях разработки и внедрения [Харитонова, Савина, Паньини, 2021: 508]. Данное требование корреспондирует с подходом Voria и соавторов, предложивших каталог fairness-aware практик, привязанных к конкретным этапам инженерного цикла [Voria et al., 2024]. Принципы правового регулирования должны включать: многоуровневость (комбинация превентивных, текущих и ретроспективных мер); документирование по жизненному циклу (обеспечение прослеживаемости решений на каждом этапе); социотехническую оценку последствий (учёт институционального контекста, правового статуса затрагиваемых решений, уязвимости затронутых групп); процессуальные гарантии на каждом уровне каскада.

На уровне постановки задачи и сбора данных правовые требования должны включать обязательную оценку воздействия на права человека (human rights impact assessment) до начала разработки высокорисковых систем. Эта оценка призвана выявить потенциальные каскадные риски ещё на этапе проектирования. Стандарты репрезентативности обучающих данных должны устанавливать минимальные требования к представленности защищаемых групп. Процедуры проверки источников данных на систематические смещения должны быть обязательными для высокорисковых систем. Документирование решений о данных (datasheets) обеспечит прослеживаемость и возможность последующего аудита.

На уровне разработки модели необходимы обязательные «fairness-ограничения» для высокорисковых систем, включая тестирование на дискриминацию по защищаемым признакам и использование метрик групповой справедливости наряду с метриками общей точности. Стандарты аудита должны предусматривать тестирование по группам, анализ различий в частоте и распределении ошибок, проверку на использование прокси-переменных, кодирующих защищаемые основания. Holstein и соавторы показали, что практики индустрии в области справедливости существенно различаются, и многие организации не имеют систематических процедур оценки дискриминационных рисков [Holstein et al., 2019: 1-16]. Это обосновывает необходимость нормативного закрепления соответствующих стандартов.

На уровне внедрения правовые механизмы должны обеспечивать оценку соответствия системы институциональному окружению и контексту применения. Обязательный «human-in-the-loop» (человек в контуре) для высокорисковых сфер должен быть детально регламентирован, чтобы исключить формальное соблюдение без реального человеческого контроля. Процедуры оспаривания и правовой защиты должны включать право субъекта на доступ к информации о логике системы и право требовать человеческого пересмотра алгоритмического решения. Мониторинг по

группам - непрерывное отслеживание различий в результатах для разных групп - должен быть обязательным для высокорисковых систем с информированием регулятора о выявленных смещениях.

На уровне обратной связи - наиболее сложном для правового регулирования необходимы механизмы выявления и прерывания петель обратной связи «feedback loops». Это включает обязательный анализ изменений в результатах системы во времени, выявление самоусиливающихся паттернов и, при необходимости, принудительное переобучение или приостановку системы. Периодический пересмотр должен предусматривать максимальные сроки эксплуатации модели без переобучения, учёт изменений в правовой среде и социальном контексте. Ghai и соавторы экспериментально подтвердили, что кумулятивный эффект множественных интервенций может быть неаддитивным и контринтуитивным, что дополнительно обосновывает необходимость системного мониторинга [Ghai et al., 2022: 3875–3885]. Процедуры приостановки и декомиссии системы при выявлении систематической дискриминации должны быть чётко регламентированы.

Для российского законодателя из проведённого анализа вытекает ряд конкретных рекомендаций. Необходим комплексный федеральный закон об искусственном интеллекте, включающий риск-ориентированную классификацию ИИ-систем с учётом потенциала каскадных эффектов. Требования для высокорисковых систем должны охватывать оценку воздействия на всех этапах жизненного цикла, стандарты «data governance» и аудита, обязательное документирование решений. Должно быть установлено ясное распределение ответственности между разработчиком, оператором и пользователем ИИ-системы, исключающее «размывание» ответственности в каскаде. Процессуальные гарантии должны включать право на информацию о применении ИИ, право на объяснение алгоритмического решения, право на его оспаривание и право на человеческий пересмотр.

Специальное регулирование ИИ в контрольно-надзорной деятельности должно предусматривать: правовые основания и ограничения

алгоритмического риск-профилирования, включая определение допустимых критериев, запрет дискриминирующих прокси-переменных и требования к прозрачности методологии; обязательный аудит систем на дискриминацию, включая тестирование на воспроизведение исторических смещений, анализ различий в частоте проверок по группам поднадзорных субъектов и выявление «feedback loops»; процедурные гарантии для поднадзорных субъектов - право знать о применении ИИ в принятии решений о проверке, право получить объяснение оснований отнесения к определённой категории риска, право оспорить результаты алгоритмического профилирования.

Институциональные механизмы должны включать создание специализированного органа по надзору за ИИ-системами, наделённого полномочиями по проведению аудита, расследованию случаев дискриминации и применению санкций. Обязательные оценки воздействия для высокорисковых систем до их внедрения обеспечат превентивный характер регулирования. Реестр высокорисковых ИИ-систем с публичным доступом к результатам аудита повысит прозрачность и общественный контроль. Талапина подчёркивает, что без институциональных механизмов мониторинга и принуждения нормативные требования рискуют остаться декларативными [Талапина, 2022: 24].

ЗАКЛЮЧЕНИЕ

Концепция дискриминационного каскада демонстрирует, что алгоритмическая дискриминация - не технический дефект отдельного этапа, а системное свойство жизненного цикла ИИ-систем, при котором локальные решения на каждом уровне наслаиваются и взаимно усиливают дискриминационные эффекты. Семь выявленных уровней каскада - от исторической предвзятости до обратных связей образуют взаимосвязанную систему, в которой формально «рациональные» технические решения на каждом этапе способны порождать системную дискриминацию.

Сравнительный анализ показал, что западная доктрина предлагает формализованную жизненно-цикловую рамку и операционализацию

каскадной модели через так называемые «pipeline-aware fairness» подходы, тогда как российские исследователи сосредоточены на правовых рисках цифровизации и конституционном принципе равенства. Предложенная в настоящей работе концепция, основанная на рамке Suresh и Guttag и обогащённая российским правовым контекстом, позволяет структурировать многоуровневые источники вреда и проектировать адресные правовые меры противодействия.

Критический анализ российского регулирования выявил системные пробелы во всех компонентах противодействия каскадным эффектам: от отсутствия стандартов «data governance» до неурегулированности обратных связей. Фрагментарность правового регулирования не обеспечивает комплексного подхода к управлению рисками дискриминации. Особую озабоченность вызывает внедрение ИИ в контрольно-надзорную деятельность при отсутствии правовых гарантий против воспроизведения исторических смещений и формирования самоусиливающихся циклов дискриминации.

Многоуровневая модель правового вмешательства, основанная на рассмотренных в настоящей статье принципах, предусматривает комбинацию превентивных, текущих и ретроспективных мер на всех стадиях жизненного цикла ИИ-систем. Практическая значимость этой модели состоит в возможности адресного противодействия каскадным эффектам на каждом уровне и проектирования интервенций с учётом их кумулятивного взаимодействия.

Разработка комплексного федерального закона об искусственном интеллекте в Российской Федерации должна учитывать каскадную природу алгоритмической дискриминации и предусматривать механизмы многоуровневого вмешательства, сочетающие правовые, организационные и технические меры. Риск-ориентированная классификация ИИ-систем должна исходить из потенциала каскадных эффектов. Специальное регулирование применения ИИ в контрольно-надзорной деятельности представляется первоочередной задачей ввиду масштабности планируемого внедрения и

значимости затрагиваемых прав. Создание институциональных механизмов надзора за ИИ-системами обеспечит эффективность правового регулирования и предотвратит его формализацию. Дальнейшие исследования должны быть направлены на детализацию стандартов аудита по каждому уровню каскада, разработку методологии оценки воздействия на защищаемые группы и изучение зарубежного опыта применения «pipeline-aware» подходов в правовом регулировании.

Список источников

Русскоязычные источники

1. Глушкова, С.И. и Летунов, Е.Д. (2020) 'Развитие нового поколения прав человека в эпоху цифровизации: цифровые права', *Вестник Уральского юридического института МВД России*, 4, с. 16–28.
2. Смирнова, А.И. (2023) 'Предвзятость как проблема алгоритмов ИИ: этические аспекты', *Философия и общество*, 3, с. 118–126. doi: 10.30884/jfio/2023.03.07.
3. Талапина, Э.В. (2022) 'Обработка данных при помощи искусственного интеллекта и риски дискриминации', *Право. Журнал Высшей школы экономики*, 15(1), с. 4–27. doi: 10.17323/2072-8166.2022.1.4.27.
4. Талапина, Э.В. (2025) 'Дискриминационный потенциал алгоритмов', *Административное право и процесс*, 2, с. 55–58. doi: 10.18572/2071-1166-2025-2-55-58.
5. Харитонова, Ю.С., Савина, В.С. и Паньини, Ф. (2021) 'Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права', *Вестник Пермского университета. Юридические науки*, 3(53), с. 488–515. doi: 10.17072/1995-4190-2021-53-488-515.

Зарубежные источники

6. Barocas, S. and Selbst, A.D. (2016) 'Big Data’s Disparate Impact', *California Law Review*, 104(3), pp. 671–732.

7. Barocas, S., Hardt, M. and Narayanan, A. (2023) *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, Mass.: The MIT Press.
8. Ghai, B., Mishra, M. and Mueller, K. (2022) 'Cascaded Debiasing: Studying the Cumulative Effect of Multiple Fairness-Enhancing Interventions', in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM 2022)*. New York: ACM, pp. 3875–3885. doi: 10.1145/3511808.3557155.
9. Grozdanovski, L. (2025) 'Non-discrimination law, the GDPR, the AI act and the - now withdrawn - AI liability directive proposal offering gateways to pre-trial knowledge of algorithmic discrimination', *AI and Ethics*, 5, pp. 5039–5062. doi: 10.1007/s43681-025-00754-0.
10. Holstein, K. *et al.* (2019) 'Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?', in *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Art. 1. doi: 10.1145/3290605.3300830.
11. Kearns, M. *et al.* (2018) 'Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness', in *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, pp. 2564–2572.
12. Sambasivan, N. *et al.* (2021) 'Re-imagining Algorithmic Fairness in India and Beyond', in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 315–328. doi: 10.1145/3442188.3445896.
13. Speicher, T. *et al.* (2018) 'A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018)*. New York: ACM, pp. 2239–2248. doi: 10.1145/3219819.3220046.
14. Suresh, H. and Guttag, J. (2021) 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle', in *Proceedings of*

EAAMO ’21: ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, pp. 1–15. doi: 10.1145/3465416.3483305.

15. Veale, M. and Zuiderveen Borgesius, F. (2021) 'Demystifying the Draft EU Artificial Intelligence Act', *Computer Law Review International*, 22(4), pp. 97–112.
16. Voria, G. *et al.* (2024) *A Catalog of Fairness-Aware Practices in Machine Learning Engineering*. arXiv:2408.16683.
17. Wachter, S., Mittelstadt, B. and Floridi, L. (2017) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law*, 7(2), pp. 76–99.

References

1. Barocas, S., Hardt, M. and Narayanan, A. (2023) *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, Mass.: The MIT Press.
2. Barocas, S. and Selbst, A.D. (2016) 'Big Data’s Disparate Impact', *California Law Review*, 104(3), pp. 671–732.
3. Ghai, B., Mishra, M. and Mueller, K. (2022) 'Cascaded Debiasing: Studying the Cumulative Effect of Multiple Fairness-Enhancing Interventions', in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM 2022)*. New York: ACM, pp. 3875–3885. doi: 10.1145/3511808.3557155.
4. Glushkova, S.I. and Letunov, E.D. (2020) 'Development of a new generation of human rights in the era of digitalization: digital rights', *Bulletin of the Ural Law Institute of the Ministry of Internal Affairs of Russia*, 4, pp. 16–28. (In Russian).
5. Grozdanovski, L. (2025) 'Non-discrimination law, the GDPR, the AI act and the - now withdrawn - AI liability directive proposal offering gateways to pre-trial knowledge of algorithmic discrimination', *AI and Ethics*, 5, pp. 5039–5062. doi: 10.1007/s43681-025-00754-0.

6. Kharitonova, Yu.S., Savina, V.S. and Pagnini, F. (2021) 'Bias of Artificial Intelligence Algorithms: Issues of Ethics and Law', *Perm University Herald. Juridical Sciences*, 3(53), pp. 488–515. doi: 10.17072/1995-4190-2021-53-488-515. (In Russian).
7. Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M. and Wallach, H. (2019) 'Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?', in *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Art. 1. doi: 10.1145/3290605.3300830.
8. Kearns, M., Neel, S., Roth, A. and Wu, Z.S. (2018) 'Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness', in *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, pp. 2564–2572.
9. Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T. and Prabhakaran, V. (2021) 'Re-imagining Algorithmic Fairness in India and Beyond', in *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 315–328. doi: 10.1145/3442188.3445896.
10. Smirnova, A.I. (2023) 'Bias as a problem of AI algorithms: ethical aspects', *Philosophy and Society*, 3, pp. 118–126. doi: 10.30884/jfio/2023.03.07. (In Russian).
11. Speicher, T., Heidari, H., Grgić-Hlača, N., Gummadi, K.P., Singla, A., Weller, A. and Zafar, M.B. (2018) 'A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018)*. New York: ACM, pp. 2239–2248. doi: 10.1145/3219819.3220046.
12. Suresh, H. and Guttag, J. (2021) 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle', in *Proceedings of EAAMO ’21: ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–15. doi: 10.1145/3465416.3483305.

13. Talapina, E.V. (2022) 'Data Processing by Means of Artificial Intelligence and the Risks of Discrimination', *Law. Journal of the Higher School of Economics*, 15(1), pp. 4–27. doi: 10.17323/2072-8166.2022.1.4.27. (In Russian).
14. Talapina, E.V. (2025) 'Discriminatory potential of algorithms', *Administrative Law and Procedure*, 2, pp. 55–58. doi: 10.18572/2071-1166-2025-2-55-58. (In Russian).
15. Veale, M. and Zuiderveen Borgesius, F. (2021) 'Demystifying the Draft EU Artificial Intelligence Act', *Computer Law Review International*, 22(4), pp. 97–112.
16. Voria, G., Sellitto, G., Ferrara, C., Abate, F., De Lucia, A., Ferrucci, F., Catolino, G. and Palomba, F. (2024) *A Catalog of Fairness-Aware Practices in Machine Learning Engineering*. arXiv:2408.16683.
17. Wachter, S., Mittelstadt, B. and Floridi, L. (2017) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law*, 7(2), pp. 76–99.