

ТЕХНИКА ВЫЯВЛЕНИЯ СОЦИАЛЬНЫХ ПОСЛЕДСТВИЙ КАК НОВЫЙ ЭТИЧЕСКИЙ СТАНДАРТ РАБОТЫ С ДАННЫМИ

*Лемешко Ю.А., студентка группы СГНЗ-42Б
Московский государственный технический университет им. Н.Э. Баумана*

*Научный руководитель: Ламинина О. Г., кандидат философских наук,
доцент кафедры «Информационная аналитика и политические технологии»*

Аннотация: В настоящее время работа с данными переживает переход от традиционных методов контроля данных (compliance) к полноценной этической ответственности. Статья обосновывает необходимость внедрения Техники выявления социальных последствий (ТВСП) как нового отраслевого стандарта. Показано, как системное использование ТВСП позволяет предотвращать цифровой редлайнинг, минимизировать социальные риски и повышать доверие общества.

Ключевые слова: этический стандарт, социальные последствия, ТВСП, алгоритмическая справедливость, цифровой редлайнинг, алгоритмическая дискриминация, этический аудит данных.

К 2026 году алгоритмические системы принимают решения, непосредственно влияющие на доступ граждан к кредитам, трудоустройству, медицинским и образовательным услугам. В этих условиях отрасль переживает парадигмальный сдвиг: формального соответствия регуляторным требованиям (compliance) становится недостаточно. На смену формальному соблюдению нормативных рамок приходит ответственная практика, ориентированная на заблаговременное выявление и нейтрализацию системных рисков [11].

Одним из наиболее острых проявлений таких рисков является цифровой редлайнинг – современная форма алгоритмической дискриминации, при которой технологии сбора и анализа данных используются для косвенного ограничения доступа определённых групп населения к финансовым, жилищным или информационным услугам. Термин восходит к исторической практике «редлайнинга», распространённой в США в 1930–1960-х годах, когда банки и страховые компании буквально очерчивали красными линиями на картах районы, населённые этническими меньшинствами или малоимущими гражданами, отказывая им в ипотеке и страховании. В цифровую эпоху этот механизм стал скрытым: вместо открытого ограничения по месту жительства или происхождению алгоритмы используют косвенные показатели (почтовые индексы, тип устройства, историю онлайн-активности, социальные связи). Формально запрещённые признаки не используются, но поскольку эти

показатели тесно связаны с реальным положением человека, система всё равно принимает решения, ограничивающие доступ отдельных групп к услугам. Ключевые механизмы цифрового редлайнинга включают: обучение моделей на исторически смещённых данных, отсутствие обязательного аудита признаков на предмет косвенной дискриминации, а также самоусиливающиеся циклы обратной связи, при которых первоначальный алгоритмический отказ ухудшает экономическое или социальное положение группы, что в дальнейшем интерпретируется моделью как подтверждение «высокого риска».

Наглядным примером служит практика таргетированной рекламы в жилищной и трудовой сферах. В ходе регуляторных проверок и независимых аудитов неоднократно выявлялось, что алгоритмы показа вакансий и объявлений об аренде жилья автоматически ограничивали показ определённым группам пользователей по возрасту, полу или геолокации. Формально системы не использовали запрещённые признаки напрямую, но опирались на историю онлайн-активности, тип устройства и паттерны потребления контента, которые статистически коррелировали с социально-экономическим статусом и демографией. В результате отдельные категории граждан системно не получали информацию о доступных возможностях, что фактически воспроизводило логику исторического редлайнинга в цифровой среде [7], [10].

Для системного анализа описанной проблематики необходимо чётко определить ключевые понятия. Этический стандарт работы с данными – это система принципов и практик, выходящая за рамки юридических обязательств и направленная на предотвращение вреда, обеспечение справедливости и уважение человеческого достоинства на всех этапах жизненного цикла данных [8], [9]. В современных условиях происходит переход от универсальных норм к контекстуальным этическим решениям, требующим предварительной оценки последствий.

Социальные последствия – это прямые и косвенные эффекты обработки данных и применения алгоритмов на положение отдельных групп населения: усиление неравенства, исключение из доступа к услугам, стигматизация или нарушение прав. Техника выявления социальных последствий (ТВСП) – это системный метод, позволяющий заранее выявлять, оценивать и минимизировать такие эффекты на этапе проектирования алгоритмов [12]. Именно данная техника становится новым этическим стандартом, превращая абстрактные ценности в конкретные рабочие процедуры.

Традиционный подход, ориентированный исключительно на соответствие законодательству, превращается в современную форму дискриминации, при которой технологии и анализ больших данных используются для исключения определённых групп населения (по региональному,

социально-экономическому или возрастному признаку) из доступа к работе, услугам, кредитами или информации [7].

Эволюция этических норм в сфере данных прошла несколько этапов. В 2010-е годы акцент стоял на приватности и защите персональных данных. К 2020–2023 годам происходило закрепление принципов справедливости, подотчётности и прозрачности [9]. В 2024–2026 годах происходит переход к проактивной оценке социального воздействия алгоритмов [10].

В России данный переход пока закреплён не единым стандартом, а отдельными нормами и стратегическими документами. Федеральный закон «О персональных данных» запрещает принятие решений, основанных исключительно на автоматизированной обработке персональных данных, если такие решения порождают юридические последствия или иным образом затрагивают права и законные интересы субъекта, а также обязывает оператора разъяснить логику решения и предоставлять возможность возражения [1], [2]. Дополнительно развитие этой сферы поддерживается Национальной стратегией развития искусственного интеллекта до 2030 года и Концепцией регулирования технологий ИИ и робототехники, что показывает формирование более широкого подхода к управлению рисками алгоритмических систем [3]–[6].

Растущее давление общества, средств массовой информации и клиентов делает неизбежным переход от формулировки «соответствуем закону» к принципу «не причиняем социального вреда» [10].

Техника выявления социальных последствий – это системный метод, применяемый на всех этапах жизненного цикла данных: от сбора информации до сопровождения готового продукта. Она состоит из пяти последовательных шагов:

1. Картирование заинтересованных сторон. Выявление всех групп, на которые может повлиять алгоритмическое решение: прямые пользователи, косвенно затронутые сообщества, регионы, уязвимые категории (малообеспеченные, пожилые, жители удалённых территорий).

2. Моделирование сценариев «что если». Создание гипотез о возможных последствиях в разных контекстах, включая краевые случаи: как поведёт себя модель при изменении экономической конъюнктуры, миграционных потоков или при появлении обратной связи, закрепляющей первоначальный отказ.

3. Оценка вероятности и силы воздействия. Количественная и качественная оценка выявленных последствий по шкале «вероятность реализации × масштаб социального вреда».

4. Проверка на этические «красные флаги». Сверка сценариев с утверждённым перечнем критериев: использование косвенных признаков, коррелирующих с защищёнными характеристиками, непропорцио-

нальное влияние на отдельные группы, отсутствие механизма обжалования автоматического решения. Применяются метрики справедливости (например, равенство долей положительных решений между группами, равенство ошибок классификации) и проводится независимая проверка выборки на репрезентативность [9].

5. Разработка корректирующих мер. Создание технических, организационных и бизнес-решений для устранения или минимизации выявленных рисков: перебалансировка данных, введение человеческого контроля для критических решений, ограничение области применения модели, публикация отчётов о проверке или полный отказ от развёртывания.

Постоянное применение данной методики необходимо, поскольку социальные последствия часто проявляются не сразу и не фиксируются на уровне классических метрик точности или конверсии. Без неё даже самые точные модели могут наносить ущерб, который впоследствии сложно и дорого исправить [11].

На практике это выглядит следующим образом. Так, например, в кредитном скоринге алгоритмы часто используют адрес проживания и историю цифровых транзакций как маркеры финансовой надёжности. Проверка показала, что жители небольших городов автоматически получали заниженные рейтинги, потому что модель ошибочно связывала их место жительства с высоким риском невозврата, игнорируя реальные доходы. После применения методики эти «адресные» признаки заменили на усреднённые экономические показатели регионов. В результате точность прогнозов не снизилась, а количество несправедливых отказов сократилось. В таргетированной рекламе внедрение этических фильтров помогло избежать показа неуместных предложений пользователям, переживающим сложные жизненные ситуации, сместив приоритет с исключительно коммерческих целей на уважение к приватности [10].

Отказ от проактивной оценки создаёт риски на разных уровнях. В масштабах страны это ведёт к закреплению цифрового неравенства: алгоритмы, обученные на исторических данных без проверки, могут ограничивать возможности уже уязвимых групп, создавая замкнутый круг [7]. Для бизнеса это риск потери доверия: пользователи быстро замечают несправедливые алгоритмические решения, а публичное обсуждение таких случаев может привести к оттоку клиентов [10]. С точки зрения регуляторов этические упущения теперь означают прямые финансовые последствия – обязательные проверки, штрафы, ограничения на работу систем. Поэтому игнорировать методику становится экономически невыгодно.

К 2026 году подход перестал быть экспериментальным и вошёл в практику многих компаний. Организации внедряют его по-разному: соз-

дают комитеты по этике данных, назначают ответственных за аудит алгоритмов, формируют рабочие группы из специалистов по машинному обучению, юристов и социологов [10]. Появилась и новая профессия – Ethical Data Analyst (аналитик по этической валидации), который отвечает за оценку рисков и контроль справедливости моделей в реальной работе [11].

Компании, которые начали внедрять методику раньше, уже видят преимущества: быстрее проходят проверки регуляторов, получают меньше жалоб, укрепляют доверие клиентов. Кроме того, прозрачная этическая политика помогает привлекать специалистов, для которых важна социальная ответственность работодателя. Так методика превращается из инструмента снижения рисков в фактор долгосрочного развития бизнеса [10].

Таким образом, техника выявления социальных последствий становится новым стандартом работы с данными. Она превращает общие этические правила в понятные шаги, которые помогают командам заранее находить и устранять риски. Это соответствует как новым требованиям закона, так и ожиданиям людей, которые хотят справедливой и прозрачной работы алгоритмов. В будущем методика станет обязательной частью разработки: появятся готовые инструменты для автоматической проверки, а сами процедуры войдут в отраслевые нормы [6], [8].

Литература и источники:

1. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» (в ред. от 2024 г.).
2. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных», ст. 16 «Права субъектов персональных данных при принятии решений на основании исключительно автоматизированной обработки их персональных данных».
3. Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации».
4. Указ Президента Российской Федерации от 15.02.2024 № 124 «О внесении изменений в Указ Президента Российской Федерации от 10.10.2019 № 490».
5. Национальная стратегия развития искусственного интеллекта на период до 2030 года (утв. Указом Президента РФ от 10.10.2019 № 490).
6. Распоряжение Правительства Российской Федерации от 19.08.2020 № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года».
7. Стариков Д.А. Этические принципы и регулятивные механизмы использования больших данных в стратегическом менеджменте // КиберЛенинка. 2025, Том 17. № s4.
8. ГОСТ Р 70.117-2024. Искусственный интеллект. Этические аспекты разработки и применения. М.: Росстандарт, 2024.
9. РАЭК. Этика данных в практике российских компаний: аналитический отчет 2024–2025. М.: РАЭК, 2025. 64 с.

10. Смирнов И.В., Козлова Е.П. От compliance к ответственности: эволюция подходов к этике данных // Вопросы информатизации. 2024. № 2. С. 112–125.
11. Трансформация этической матрицы в цифровую эпоху: материалы всероссийской научной конференции. М.: Социологический факультет МГУ, 2022. 184 с.
12. Mehrabi N., Morstatter F., Saxena N. et al. A Survey on Bias and Fairness in Machine Learning // ACM Computing Surveys. 2021. Vol. 54, № 6. Art. 115.
13. Противодействие фальсификации истории великой отечественной войны / Бочарников И.В., Суздалева Т.Р., Федоров К.В., Криворучко А.А., Петренко А.И., Зеленков М.Ю., Кандыбович С.Л., Разина Т.В., Овсянникова О.А., Трипольский В.Б. Москва, 2020.
14. Ремарчук В.Н. Информационно-аналитическая деятельность: проблемы и перспективы // Вестник Академии военных наук. 2023. № 1 (82). С. 31–35.