

Мир науки. Педагогика и психология / World of Science. Pedagogy and psychology <https://mir-nauki.com>

2025, Том 13, № 2 / 2025, Vol. 13, Iss. 2 <https://mir-nauki.com/issue-2-2025.html>

URL статьи: <https://mir-nauki.com/PDF/95PSMN225.pdf>

5.3.3. Психология труда, инженерная психология, когнитивная эргономика (психологические науки)

Ссылка для цитирования этой статьи:

Камальдинова, З. Ф. Проблема создания этических и прозрачных алгоритмов для объяснимости решений в автоматизированных системах рекрутмента с учетом требований регулирования конфиденциальности данных / З. Ф. Камальдинова, В. С. Панарин, Р. С. Чуб // Мир науки. Педагогика и психология. — 2025. — Т. 13. — № 2. — URL: <https://mir-nauki.com/PDF/95PSMN225.pdf>.

For citation:

Zulfiya F.K., Panarin V.S., Chub R.S. The problem of creating ethical and transparent algorithms to ensure explainability of decisions in automated recruitment systems, while complying with data privacy regulations. *World of Science. Pedagogy and psychology*. 2025;13(2): 95PSMN225. Available at: <https://mir-nauki.com/PDF/95PSMN225.pdf>. (In Russ., abstract in Eng.).

УДК 159.99

Камальдинова Зульфия Фаисовна

ФГБОУ ВО «Самарский государственный технический университет», Самара, Россия
Доцент кафедры «Информатика и вычислительная техника»

Кандидат технических наук, доцент

E-mail: kamal_zzz@mail.ru

ORCID: <https://orcid.org/0000-0002-9999-5436>

РИНЦ: https://elibrary.ru/author_profile.asp?id=767354

Панарин Владимир Сергеевич

ФГБОУ ВО «Самарский государственный технический университет», Самара, Россия

E-mail: panvlad2@mail.ru

ORCID: <https://orcid.org/0009-0002-4173-517X>

Чуб Роман Сандипович

ФГБОУ ВО «Самарский государственный медицинский университет» Минздрава России, Самара, Россия

E-mail: romachub1@gmail.com

ORCID: <https://orcid.org/0009-0004-6785-3048>

**Проблема создания этических
и прозрачных алгоритмов для объяснимости
решений в автоматизированных системах рекрутмента
с учетом требований регулирования
конфиденциальности данных**

Аннотация. В данной статье исследуется развитие этических и прозрачных алгоритмов в автоматизированных системах рекрутмента. Рассматриваются этапы эволюции этих систем, анализируются требования и их влияние на проектирование алгоритмов. В работе приведены примеры использования. Освещаются вопросы объяснимости и прозрачности в автоматизированных решениях, приводятся примеры как можно добиться этого различными способами и методами. Также обсуждаются методы создания этических алгоритмов и приводятся влияния реальных примеров успешных кейсов их внедрения. В дополнении к этому также приводятся контрпримеры иллюстрирующие то, как алгоритмы могут негативно влиять на процесс найма, дискриминировать отдельные группы лиц. Приведены причины данной предвзятости искусственного интеллекта.

В связи с этим статье также отмечена дискуссия в сторону ограничения применения автоматизированных систем на основе искусственного интеллекта. Кроме того, в статье приводятся примеры того, как такие автоматизированные системы влияют на бизнес: позволяют удерживать сотрудников, улучшает их лояльность и повышает конкурентное преимущество компаний.

В статье подчеркивается значимость междисциплинарного подхода и сотрудничества с экспертами в области права, этики и управления человеческими ресурсами. При этом отдельное внимание уделяется необходимости соответствия нормативным требованиям, таким как «Общий регламент по защите данных», которые направлены на обеспечение прозрачности и объяснимости автоматизированных систем принятия решений.

Прогнозируются будущие направления развития, отмечаются перспективы, включая применение новых технологий, для повышения точности и справедливости решений.

Ключевые слова: машинное обучение; автоматизированные системы рекрутмента; этические алгоритмы; алгоритмическая предвзятость; прозрачность; объяснимость; регуляторные требования; междисциплинарный подход

Введение

Современное развитие технологий обучения машин и автоматизации существенно повлияло на рекрутинг, превратив его из утомительных ручных процессов в высокоэффективные автоматизированные системы. Эти системы, применяющие мощные алгоритмы для обработки и анализа больших объемов данных, кардинально изменили методы принятия решений в найме. Как известно, цикл работы с сотрудником крайне субъективен. И хотя искусственный интеллект может повысить эффективность каждого из процессов, повышение точности и устранение предвзятости остаются основными вызовами для этих систем [1].

Данная работа исследует комплексный подход к проектированию таких систем, выделяя аспекты прозрачности, справедливости, регулирующего воздействия и практической реализации.

Обзор использования нейросетей при отборе персонала

Подбор — один из самых трудоемких процессов во всем цикле работы с сотрудником, так как на одну вакансию необходимо рассмотреть десяток человек, выбрать двух-трех, со всеми провести интервью, предоставить обратную связь. И все это помноженное на субъективность рекрутера. При этом в рекрутменте самое большое количество рутинных задач, которые могут быть автоматизированы искусственным интеллектом. Особенно это касается массового подбора.¹

Стартапы по типу Textio оценивают описание вакансий по ряду критериев:

- привлекательность для кандидатов;
- отсутствие дискриминации по полу или расе;
- уровень сложности и длина текста.

¹ Финкельштейн, Г., Шатров Ю. Как искусственный интеллект меняет HR // HRTimes № 35 Апрель 2020 // [Электронный ресурс]. URL: <https://www.ecopsy.ru/insights/kak-iskusstvennyy-intellekt-menyaet-hr/> (дата обращения: 18.03.2025).

На основе оценки выдаются рекомендации по улучшению вакансии.

Отдельный большой пласт технологий посвящен поиску кандидатов и оценке уровня их соответствия требованиям к квалификации. Множество стартапов ищут в социальных сетях и профессиональных сообществах «пассивных» кандидатов, которые могут быть эффективны на определенных позициях. Самые известные стартапы: Blendoor, Ideal, Harver, HiringSolved, Headstart.

Чат-боты, основанные на искусственном интеллекте, способны вести диалог с кандидатом и далее выставлять оценку его потенциальной эффективности. Важно, что чат-боты с искусственным интеллектом в отличие от обычных не требуют прямого программирования правил, каким образом необходимо отвечать на вопросы человека. Наоборот, искусственный интеллект способен проинтерпретировать даже не прямой ответ соискателя. За рубежом чат-боты разрабатывают компании Mya Systems, Paradox, Text Recruit. В России — Яндекс.Таланты.

Историческое развитие автоматизированных систем рекрутмента свидетельствует о значительных изменениях. От элементарных инструментов до интегрированных систем с машинным обучением. Первоначальные приложения автоматизировали базовые задачи, такие как сортировка резюме и фильтрация кандидатов по ключевым словам. Однако современные алгоритмы используют глубокое машинное обучение для более глубокой оценки потенциала кандидатов [2].

Исследования в области алгоритмической предвзятости подчеркивают растущую озабоченность недостаточной объяснимостью решений, принимаемых этими системами [3]. Модели часто действуют по принципу «черного ящика», в результате чего конечные пользователи не могут отчетливо проследить логику, ведущую к конкретным решениям. Это вызывает обеспокоенность среди клиентов и работодателей, что ведет к детализированным исследованиям по улучшению интерпретируемости и прозрачности алгоритмов [4].

Существует много подходов к решению заявленных проблем, включая разработку гибридных моделей, которые объединяют машинное обучение с классическими статистическими методами для достижения лучшей объяснимости и устойчивости [5]. Немаловажно учесть, что интеграция методов интерпретируемости в карьерные системы улучшает доверие к ним. Искусственный интеллект (ИИ) нуждается в переосмыслении его проектирования, включающем не только технологические, но и социальные, и этические ориентиры [6].

Субъективность алгоритмов автоматического рекрутмента

Несколько лет назад исследовательницы из Лондонской школы экономики с помощью эмпирических методов изучили, как работают алгоритмы таргетированной рекламы Facebook² и Google. Их целью было выяснить, кому чаще показывают объявления о высокооплачиваемых вакансиях в традиционно «мужских» сферах — науке, технологиях, инженерии и математике (STEM). В результате они опубликовали в 2018 году статью «Алгоритм предвзятости? Эмпирическое исследование явной дискриминации по признаку пола в рекламе вакансий STEM» [7].

Исследование показало, что несмотря на то, что сами объявления были гендерно нейтральными, алгоритмы Facebook² и Google демонстрировали их мужчинам значительно чаще, чем женщинам. Однако причина этого крылась не в преднамеренной дискриминации, а в самой логике работы алгоритмов. Системы машинного обучения, оптимизируя расходы на

² Социальная сеть признана экстремистской организацией запрещена на территории Российской Федерации.

рекламу, «решили», что размещение объявлений для мужчин экономически выгоднее, так как их чаще нанимают на подобные должности.

Также случилось с рекрутинговой системой Amazon: оказалось, что она нарочно отметала женские резюме, которые соискательницы подавали на ИТ-должности. Компании, которая разработала алгоритм для поиска лучших кадров с 2014 года, уже в 2017 году пришлось распустить команду разработчиков и признать проект неудачным.³

Согласно отчёту Гарвардской школы бизнеса, автоматизированные системы фильтрации резюме в США привели к исключению миллионов подходящих кандидатов. Около 88 % работодателей отметили, что высококвалифицированные соискатели часто исключаются алгоритмами из-за несоответствия точным критериям, даже если они обладают необходимыми навыками и опытом.⁴

При этом алгоритм сам по себе не может быть ни этичным, ни дискриминирующим. Искусственный интеллект не пытается воспроизвести мышление человека, а просто решает поставленную перед ним задачу на основе данных, на которых был обучен. В случае с рекламой вакансий проблема заключалась не в том, *как* алгоритм принимал решение, а в *исходных данных*, которые формировали его логику.

В 2019 году MIT Review опубликовал статью «This is how AI bias really happens — and why it’s so hard to fix». По мнению автора, у «предвзятости» искусственного интеллекта (ИИ) есть три причины:

1. Размытая задача. Например, скоринговому алгоритму банка нужно определить кредитные качества клиента. Четкого определения этому понятию нет.
2. Процесс сбора данных. Данные могут быть либо нерепрезентативными, либо уже быть необъективными.
3. Процесс подготовки данных. Предрассудки программистов могут сказаться при выборе атрибутов для алгоритма, с помощью которых он будет оценивать соискателя вакансии.⁵

Регуляторные требования и их влияние на алгоритмы

На фоне растущих правовых требований, таких как «Общий регламент по защите данных» (GDPR), компании сталкиваются с необходимостью адаптации своих алгоритмических решений [3]. GDPR и другие законодательные акты направлены на обеспечение прозрачности и объяснимости автоматизированных систем принятия решений. Это предполагает, что компании должны не только объяснять решения, которые принимаются алгоритмами, но и гарантировать защиту персональных данных, используемых в таких процессах.

Внедрение законов связано с набором ключевых аспектов, включая предотвращение дискриминации и обеспечение прозрачности для конечных пользователей [8]. Типичным вызовом здесь является необходимость в проведении внутренних аудитов и управления

³ Amazon scraps secret AI recruiting tool that showed bias against women [Электронный ресурс]. URL: <https://www.reuters.com/article/us-amazon-com-jobsautomation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-biasagainst-women-idUSKCN1MK08G> (дата обращения: 28.03.2025).

⁴ Алгоритмы подбора кадров ошибочно отвергли миллионы кандидатов в США [Электронный ресурс]. URL: <https://trends.rbc.ru/trends/education/613725fe9a7947812c055498> (дата обращения: 28.03.2025).

⁵ This is how AI bias really happens — and why it’s so hard to fix // MIT Review [Электронный ресурс]. URL: <https://clck.ru/NNQKs> (дата обращения: 18.03.2025).

качеством алгоритмов [9]. Результирующая необходимость требует от компаний весомых инвестиций в улучшение систем управления рисками и мониторинга качества алгоритмов.

Компании должны подходить к проектированию искусственного интеллекта с акцентом на безопасность, конфиденциальность и соответствие требованиям на каждом этапе разработки системы. Такой подход ведёт не только к более этически-приемлемым алгоритмам, но также формирует доверие между рекрутерами, соискателями на работу и компаниями-разработчиками [10].

В тоже время необходимо отметить, что дискуссия развивается также в сторону ограничения применения автоматизированных систем на основе ИИ. Юлия Харитонова предполагает, что прозрачность и объяснимость технологии искусственного интеллекта важна не только для защиты персональных данных, но и в иных ситуациях автоматизированной обработки данных, когда для принятия решений недостающие из входящей информации технологические данные восполняются из открытых источников. Законодательство может лишь устанавливать определённый стандарт, представляющий собой компромисс между точностью и интерпретируемостью их результатов, а также правами участников общественных отношений. Введение обязательной сертификации моделей искусственного интеллекта, используемых в критически важных сферах, позволит урегулировать вопросы ответственности субъектов, обязанных применять такие системы. В отношении профессиональной ответственности специалистов, таких как медицинские работники, военные, а также представители корпоративного управления, необходимо ограничить обязательное применение искусственного интеллекта в тех случаях, когда его прозрачность не является достаточной. Кроме того, правовая дискуссия в данной сфере должна быть направлена на разработку норм, обеспечивающих право на отказ от использования автоматизированных систем при принятии решений, а также право оспаривать решения, принятые на основе алгоритмической обработки данных [11].

Объяснимость в автоматизированных системах

Объяснимость алгоритмов представляет собой основополагающий принцип, поддержку которого требует современное развитие информационных технологий [12]. Возможность проследить решения по всей цепочке их принятия является критичным для пользователей и заказчиков. Поднимается вопрос о том, каким образом конкретные данные могут оказывать влияние на решения системы, что помогает избежать чувства недоверия к технологии.

Одним из эффективных подходов являются разработки объяснительных моделей, таких как Local Interpretable Model-agnostic Explanations (LIME) и SHapley Additive exPlanations (SHAP) [5].

Эти подходы делают возможным глубокое понимание влияния индивидуальных данных на принятие решений алгоритмами. Такое понимание имеет критически важное значение для устранения ошибок и неправильной интерпретации данных, тем самым снижая потенциальные риски при внедрении решений [8].

Для достижения высокого уровня объяснимости, разработчикам необходимо пересматривать подход при проектировании алгоритмов, смещая акценты на прозрачность и доступность информации для конечных пользователей [1].

Достижение этого состояния может быть основой для соблюдения современных нормативных требований и устранения субъективности. По мере увеличения внимания к вопросу объяснимости, возрастают и её значимость, и потенциал приложений для разработки социально-ответственного машинного обучения [3].

Создание этических алгоритмов

Согласно статье «*Ethics of Artificial Intelligence and Robotics*», этические алгоритмы должны следовать нормативным принципам, которые включают в себя предотвращение предвзятости, защиту прав человека и соблюдение правовых стандартов [13]. Концепция «privacy by design» диктует обязательное включение защиты данных на всех этапах проектирования и внедрения систем. Это позволяет минимизировать риски утечки данных и нарушения конфиденциальности.

В работе «*A Survey on Bias and Fairness in Machine Learning*» описывается проблема предвзятости в алгоритмах машинного обучения и предлагаются методы её устранения, включая алгоритмическую проверку и механизмы корректировки данных, используемых в обучении [14]. Предвзятость при разработке алгоритма может возникнуть из-за ошибок в данных, и недостаточной представленности отдельных социальных групп [5]. Дополнительно для устранения ошибок применяются корректирующие меры, такие как предобработка данных и адаптивное обучение моделей [8]. Такие стратегии создания справедливых алгоритмов обеспечивают количественные и качественные механизмы для контроля предвзятого поведения систем [2].

Успешная интеграция этических алгоритмов требует учета межкультурных различий и разнообразия данных, предполагая активное использование знаний из этнологии и социологии. Такое междисциплинарное обобщение помогает в понимании и адаптации автоматизированных решений, чтобы они соответствовали социальным нормам и ожиданиям [9].

При этом в «*Principles alone cannot guarantee ethical AI*» отмечается, что принципы этичности ИИ схожи с принципами этики в медицинской сфере. Но подходы совершенно отличаются так, как у ИИ нет единой цели, что затрудняет этическое принятие решений, нет устоявшейся профессиональной этики, и нет строгих стандартов контроля. Автор предлагает определить долгосрочные цели и подходы к этике, учитывающие специфику ИИ, внедрить лицензирование разработчиков ИИ в высокорискованных сферах. Отдельно автор отмечает важность того, что этические нормы должны исходить от самого сектора, создавая устойчивую базу [16].

Методы создания объяснимых решений

Интерпретируемое машинное обучение предоставляет различные методики, направленные на повышение понимания функционирования алгоритмов. Примером таких мер являются модели на основе правил, а также визуализационные инструменты, что позволяет выявлять значимость различных параметров [5]. Поточковые графики и тепловые карты — это примеры инструментов, которые могут помочь разработчикам и пользователям определить критически важные факторы, влияющие на конечные результаты. Существуют методы оценки важности признаков: Эти методы определяют, какие входные признаки наиболее влияют на предсказания модели. Например, метод Permutation Feature Importance (PFI) оценивает влияние каждого признака путём случайного перемешивания его значений и анализа изменения точности модели. Чем сильнее изменение, тем более значимым считается признак. Однако PFI объясняет общее поведение модели и не предоставляет интерпретации для отдельных предсказаний.

Методы, такие как SHAP (SHapley Additive exPlanations), используют концепцию значений Шепли из теории кооперативных игр для оценки вклада каждого признака в предсказание модели. SHAP обеспечивает как глобальные, так и локальные объяснения, позволяя понять влияние каждого признака на уровне отдельных предсказаний [6; 17; 18].

Основная идея заключается в том, чтобы создавать алгоритмы, которые поддаются оценке и анализу, позволяя конечным пользователям не только понимать, как осуществляется процесс принятия решений, но и корректировать деятельность алгоритмов на основе собранных данных [4]. Эти меры направлены на снижение недоверия и на повышение точности работы систем.

Социальные и экономические аспекты внедрения объяснимых алгоритмов играют важную роль в создании доверия к автоматизированным системам. Это, в свою очередь, способствует снижению потенциальных рисков судебных разбирательств, улучшает общественное восприятие бренда, что повышает конкурентоспособность компании на рынке [3].

Психологические аспекты и влияние

Практика внедрения объяснимых и этичных алгоритмов в рекрутинговые системы демонстрирует множество успешных примеров, где соблюдение принципов прозрачности и справедливости приводит к улучшению процессов найма [13; 19].

Компании, принявшие модели с высокой степенью объяснимости, сообщили о снижении числа ошибок и уровня предвзятости, что укрепило доверия со стороны клиентов и пользователей. В исследовании, проведенном LinkedIn, 76 % сотрудников, нанятых с использованием более этичных алгоритмов, оставались в компании дольше двух лет, по сравнению с 60 % тех, кто прошел по традиционным методам.

Существуют также примеры, когда отсутствие прозрачности и непредвзятости привели к негативным последствиям [10; 19]. Изучение таких случаев позволяет выявить критические недостатки и разработать стратегии для их устранения, способствуя совершенствованию практик внедрения [4].

Обзор решений на рынке показывает, что многие платформы вкладывают значительные усилия в объяснимость и справедливость своих алгоритмов, предлагая инструменты аудита и корректировки алгоритмов для оценки их технического соответствия.

Объяснимые алгоритмы оказывают непосредственное влияние на эффективность рекрутинговых процессов, делая принимаемые решения более обоснованными и прозрачными. Это положительно сказывается на лояльности и удовлетворенности сотрудников, улучшая общее восприятие корпоративной культуры и бренда компании.

В долгосрочной перспективе компании, успешно внедряющие эти практики, формируют более устойчивые корпоративные стратегии [3]. Такие компании получают значительные конкурентные преимущества за счет соблюдения строгих стандартов прозрачности и этичности, что улучшает их имидж и рыночную долю [5]. По данным исследования McKinsey, компании с высоким уровнем разнообразия в составе сотрудников на 35 % более вероятно превышают свои финансовые цели по сравнению с менее разнообразными компаниями.⁶

Однако на пути внедрения таких систем существуют значительные барьеры, такие как необходимость значительных перераспределений ресурсов и потенциальное сопротивление изменениям со стороны персонала [19]. Рекомендуется применять поэтапный подход к адаптации, который включает комплексное обучение и вовлечение всех заинтересованных сторон в процесс изменений.

⁶ Эффективность команды связана с разнообразием её состава [Электронный ресурс]. URL: <https://businessrevisor.ru/2019/04/high-performing-teams-diversity/> (дата обращения: 28.03.2025).

Обсуждение и перспективы

Создание этических и прозрачных алгоритмов для автоматизированных систем рекрутмента требует комплексного подхода с включением экспертизы из различных дисциплин, таких как право, этика и управление человеческими ресурсами [8]. Современные исследования продолжаются, фокусируясь на углубленном изучении взаимосвязей между технологическими разработками и социальными процессами [10]. Чтобы справиться с описанными проблемами внедрения нейросетей, в будущем следует сосредоточиться на совершенствовании методов выявления и устранения предвзятости, повышении прозрачности с помощью объяснимых методов ИИ и разработке гибридных моделей, сочетающих эффективность ИИ с эмпатией человеческой мысли [20].

Будущие достижения в этом направлении должны сосредоточиться на интеграции разнообразных данных и новых технологий, для создания альтернативных методов анализа данных и принятия решений [1]. Возможности для улучшения практик в этой области велики, и их использование может привести к позитивным изменениям, как на уровне отдельных компаний, так и в индустрии в целом [12].

Заключение

В заключение стоит подчеркнуть, что создание этических и прозрачных алгоритмов для автоматизированных систем рекрутмента требует всеобъемлющего подхода, сочетающего передовые технологии и лучшие практики управления данными. Эти инициативы должны идти рука об руку с соблюдением нормативных требований, что укрепит доверие пользователей и улучшить корпоративную репутацию.

Такой подход может оказаться критически важным для компаний, стремящихся к долгосрочному успеху и устойчивости в быстро меняющемся цифровом мире. В конечном итоге, это способствует построению более справедливого общества, где технологии работают в интересах всех участников процесса социальных взаимодействий.

ЛИТЕРАТУРА

1. Jobin, A., Ienca, M., Vayena, E. Artificial Intelligence: The Global Landscape of Ethics Guidelines // Nature Machine Intelligence. 2022. URL: https://www.researchgate.net/publication/334082218_Artificial_Intelligence_the_global_landscape_of_ethics_guidelines (дата обращения: 28.03.2025).
2. Selbst, A.D., Powles, J. Meaningful Information and the Right to Explanation // Data Protection & Privacy Journal. 2023. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125 (дата обращения: 28.03.2025).
3. Sharma, V. AI, Ethics, and Society: Balancing Innovation with Responsibility // Science and Engineering Ethics. 2025. URL: https://www.researchgate.net/publication/388062868_Ethics_in_AI_Balancing_innovation_and_responsibility (дата обращения: 18.03.2025).
4. Suresh, H., Gutttag, J.V. A Framework for Understanding Unintended Consequences of Machine Learning // Communications of the ACM. 2022. URL: https://www.researchgate.net/publication/330726492_A_Framework_for_Understanding_Unintended_Consequences_of_Machine_Learning (дата обращения: 18.03.2025).

5. Shneiderman, B. Human-Centered AI: Reliable, Safe & Trustworthy Technologies // ACM Transactions on Intelligent Systems and Technology. 2023. URL: https://www.researchgate.net/publication/375743140_Human-Centered_Artificial_Intelligence_Reliable_Safe_Trustworthy (дата обращения: 18.03.2025).
6. Аванесян Г.Г. Перспективы эффективного взаимодействия искусственного интеллекта и личности человека // Мир науки. Педагогика и психология. 2024. Т. 12, № 3. URL: <https://mir-nauki.com/PDF/77PSMN324.pdf> (дата обращения: 18.03.2025).
7. Lambrecht, A., Tucker, C. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. 2019. URL: <https://www.researchgate.net/publication/331038992> (дата обращения: 18.03.2025).
8. Zafar, M.B., Valera, I., Rodriguez, M.G., Gummadi, K.P. Fairness Constraints: Mechanisms for Fair Classification // Conference on Fairness, Accountability, and Transparency (FAT*). 2022. URL: <https://arxiv.org/abs/1507.05259> (дата обращения: 18.03.2025).
9. Lepri, B., Oliver, N., Pentland, A., Venkatramanan, V. Fair, Transparent, and Accountable Algorithmic Decision-Making Processes // IEEE Internet Computing. 2022. URL: https://www.researchgate.net/publication/319127327_Fair_Transparent_and_Accountable_Algorithmic_Decision-making_Processes_The_Premise_the_Proposed_Solutions_and_the_Open_Challenges (дата обращения: 28.03.2025).
10. Bryson, J.J., Winfield, A.F.T. Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems // Proceedings of the IEEE. 2022. URL: https://www.researchgate.net/publication/316898639_Standardizing_Ethical_Design_for_Artificial_Intelligence_and_Autonomous_Systems (дата обращения: 28.03.2025).
11. Харитоновна, Ю.С. Правовые средства обеспечения принципа прозрачности искусственного интеллекта // Journal of Digital Technologies and Law. 2023. Т. 1, № 2. С. 337–358. DOI: <https://doi.org/10.21202/jdtl.2023.14>.
12. Raji, I.D., Buolamwini, J. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products // Conference on AI, Ethics, and Society. 2022. URL: https://www.researchgate.net/publication/334383981_Actionable_Auditing_Investigating_the_Impact_of_Publicly_Naming_Biased_Performance_Results_of_Commercial_AI_Products (дата обращения: 28.03.2025).
13. Vincent C. Müller. Ethics of Artificial Intelligence and Robotics // Stanford Encyclopedia of Philosophy [Электронный ресурс]. URL: https://www.researchgate.net/publication/340435326_Ethics_of_Artificial_Intelligence_and_Robotics (дата обращения: 18.03.2025).
14. Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, Aram Galstyan, A Survey on Bias and Fairness in Machine Learning // University of Southern California URL: https://www.researchgate.net/publication/340435326_Ethics_of_Artificial_Intelligence_and_Robotics (дата обращения: 18.03.2025).
15. Mittelstadt, B. Principles alone cannot guarantee ethical AI // Journal of AI and Ethics. 2023. URL: <https://arxiv.org/pdf/1906.06668> (дата обращения: 18.03.2025).

16. Mailson R.S. SHapley Additive exPlanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis // Machine Learning and Knowledge Extraction 2024 № 6(1):316–341 [Электронный ресурс]. URL: https://www.researchgate.net/publication/378063501_SHapley_Additive_exPlanations_SHAP_for_Efficient_Feature_Selection_in_Rolling_Bearing_Fault_Diagnosis (дата обращения: 28.03.2025).
17. Magham, R.K. Mitigating Bias in AI-Driven Recruitment: The Role of Explainable Machine Learning // International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2022. Т. 10, № 5. С. 461–469. URL: https://www.researchgate.net/publication/384802794_Mitigating_Bias_in_AI-Driven_Recruitment_The_Role_of_Explainable_Machine_Learning_XAI (дата обращения: 28.03.2025).
18. Mujtaba, D.F., Mahapatra, N.R. Fairness in AI-Driven Recruitment: Challenges, Metrics, Methods, and Future Directions // Information Systems Research. 2022. URL: <https://arxiv.org/html/2405.19699v2> (дата обращения: 28.03.2025).
19. Rigotti, C. Fairness, AI & recruitment // Computer Law & Security Review. 2024. URL: https://www.researchgate.net/publication/385720756_Diversity_and_Inclusion_in_AI_for_Recruitment_Lessons_from_Industry_Workshop (дата обращения: 18.03.2025).
20. Javaid, B., Pearl, J. Ethical Challenges of Artificial Intelligence Implementation in Human Resources: Implications for Job Performance Evaluation // Computers in Human Behavior Reports. 2023. URL: https://www.researchgate.net/publication/381999507_Ethical_Challenges_of_Artificial_Intelligence_Implementation_in_Human_Resources_Implications_for_Job_Performance_Evaluation (дата обращения: 28.03.2025).

Zulfiya Faisovna Kamaldinova

Samara State Technical University, Samara, Russia

E-mail: kamal_zzz@mail.ru

ORCID: <https://orcid.org/0000-0002-9999-5436>

RSCI: https://elibrary.ru/author_profile.asp?id=767354

Panarin Vladimir Sergeevich

Samara State Technical University, Samara, Russia

E-mail: panvlad2@mail.ru

ORCID: <https://orcid.org/0009-0002-4173-517X>

Chub Roman Sandipovich

Samara State Medical University, Samara, Russia

E-mail: romachub1@gmail.com

ORCID: <https://orcid.org/0009-0004-6785-3048>

The problem of creating ethical and transparent algorithms to ensure explainability of decisions in automated recruitment systems, while complying with data privacy regulations

Abstract. This article discusses the development of ethical and transparent algorithms in automated recruitment systems. It examines the stages of evolution of these systems, analyzing the requirements and impact on algorithm design. The paper provides examples of usage, highlighting the importance of explainability and transparency in automated solutions. It discusses methods for creating ethical algorithms and presents real-world examples of their successful implementation. Additionally, counterexamples are provided to illustrate potential negative impacts of algorithms on the hiring process, such as discrimination against certain groups. The article also discusses the reasons for such bias in artificial intelligence and suggests ways to limit the use of AI-based systems. In addition, the article provides examples of how automated systems can benefit businesses: they can help retain employees, increase their loyalty, and give companies a competitive advantage. The article emphasizes the importance of a collaborative approach between different professionals, such as lawyers, ethicists, and HR specialists. At the same time, it emphasizes the need to comply with regulations like General Data Protection Regulation, which aim to ensure transparency and explainability in automated decision-making. Future developments are discussed, including the potential use of new technologies to improve accuracy and fairness in decision-making processes.

Keywords: automated recruitment systems; ethical algorithms; transparency; explainability; regulatory requirements; interdisciplinary approach