Секция «Социальные технологии в управлении общественными процессами»

Проблема «черного ящика» и перспективы ответственного ИИ в управлении

Научный руководитель – Гавриленко Ольга Владимировна

Отрыванов Даниил Игоревич
Студент (бакалавр)
Московский государственный университет имени М.В.Ломоносова, Социологический факультет, Кафедра социальных технологий, Москва, Россия
E-mail: otr.devs@gmail.com

Внедрение искусственного интеллекта в HR-процессы становится повсеместной практикой в ведущих компаниях мира. Алгоритмы машинного обучения используются для отбора персонала, оценки эффективности сотрудников, прогнозирования текучести кадров и управления рабочими процессами. Согласно отчету McKinsey, 63% организаций сталкиваются с проблемой неточности прогнозов ИИ, 52% — с вопросами кибербезопасности, а 51% — с угрозой нарушения прав интеллектуальной собственности [5].

Несмотря на эффективность автоматизированных решений, ключевым вызовом остается проблема прозрачности работы алгоритмов. Современные модели машинного обучения представляют собой сложные математические конструкции, в которых невозможно точно проследить, как принимается то или иное управленческое решение, что приводит к рискам необоснованных кадровых решений, усилению социального неравенства и снижению доверия сотрудников к цифровым инструментам управления.

Цель данного исследования — рассмотреть основные этические вызовы применения ИИ в управлении персоналом, определить принципы ответственного ИИ.

Алгоритмы глубинного обучения, используемые в современных HR-системах, работают на основе многослойных нейросетей, механизм работы которых зачастую остается неизвестным даже их создателям, что затрудняет интерпретацию решений ИИ, что особенно критично при оценке сотрудников, принятии решений о найме или увольнении[3]. Так, проблема непрозрачности ИИ тесно связана с качеством данных. Согласно данным McKinsey, 70% лидеров в области ИИ сталкиваются с проблемами чистоты и достоверности данных. Если алгоритм обучается на исторических данных, в которых заложены предвзятые кадровые решения (например, предпочтение мужчин при найме в IT-секторе), система будет автоматически воспроизводить эти предвзятости. Это может приводить к дискриминации сотрудников по полу, возрасту или другим социальным характеристикам [5].

Алгоритмические HR-системы также создают новые угрозы для кибербезопасности и защиты данных сотрудников. Программы мониторинга персонала позволяют отслеживать активность сотрудников, анализировать рабочую переписку и даже измерять уровень вовлеченности через анализ поведения в корпоративных системах. Однако их чрезмерное использование может привести к цифровому тоталитаризму внутри организации, снижению доверия и усилению стрессовых факторов у сотрудников.

Кроме того, компании, использующие ИИ в управлении персоналом, сталкиваются с вызовами регулирования и внедрения стандартов ответственного ИИ. Исследования показывают, что 48% организаций рассматривают управление рисками ИИ как ключевой вызов, требующий строгого соблюдения этических стандартов [5]. Это особенно актуально в таких сферах, как стратегическое управление и логистика, где ошибки в прогнозах могут привести к сбоям в цепочках поставок и финансовым потерям.

Для минимизации рисков, связанных с применением ИИ в управлении персоналом, необходима разработка концепции ответственного ИИ, включающей следующие принципы [4]:

- Прозрачность и объяснимость, где алгоритмы должны быть понятными и интерпретируемыми, а решения ИИ — подлежащими разъяснению.
- Справедливость и отсутствие дискриминации — системы ИИ должны исключать предвзятость в отношении пола, возраста, национальности и других факторов.
- Защита персональных данных — внедрение механизмов защиты информации сотрудников, соответствующих международным стандартам.
- Человекоцентричный подход — ИИ должен дополнять, а не заменять управленческие решения, обеспечивая баланс между автоматизацией и человеческим фактором.
- Этическое регулирование и аудит — компании должны проводить регулярные проверки алгоритмов ИИ, выявляя потенциальные ошибки и предвзятости.

Например, в России концепция ответственного искусственного интеллекта (ИИ) приобретает всё большее значение, отражая стремление к этичному и прозрачному внедрению ИИ-технологий в различные сферы жизни, как на корпоративном, так и на государственном уровне. Так, при создании генеративной нейросети Cotype от МТС особое внимание уделяется обучению модели на специально подготовленных этичных датасетах, что обеспечивает соответствие ответов нейросети как интересам компании, так и общества в целом, повышая доверие и лояльность пользователей [2]. Такой подход не только способствует соблюдению законодательных норм, но и укрепляет репутацию компании на рынке. Более того, в 2024 году были внесены изменения в Федеральный закон «Об экспериментальных правовых режимах в сфере цифровых инноваций», предусматривающие ответственность за причинение вреда при использовании ИИ-решений [1].

Таким образом, искусственный интеллект становится ключевым инструментом в управлении персоналом, однако его применение сопряжено с рядом этических вызовов. Проблема «черного ящика» в алгоритмах машинного обучения, предвзятость данных и риски нарушения кибербезопасности требуют внедрения принципов ответственного ИИ, направленных на прозрачность, защиту данных и обеспечение справедливости решений. Для эффективного регулирования ИИ в HR, по нашему мнению, необходимо: разрабатывать гибридные модели управления, в которых алгоритмы дополняют, а не подменяют человеческие решения, внедрять стандарты этичного ИИ, позволяющие минимизировать дискриминационные риски, использовать аудит алгоритмов для выявления предвзятости и повышения доверия к цифровым HR-инструментам.

Источники и литература

1) Федеральный закон "О внесении изменений в Федеральный закон "Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации" от 08.07.2024 N 169-ФЗ (последняя редакция)//КонсультантПлюс. 2024 [Электронный ресурс]. -Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_480370/, свободный (27.02.2025)
2) Выгодная этика: как ответственный ИИ помогает зарабатывать больше//МТС AI. 2024 [Электронный ресурс]. -Режим доступа: https://mts.ai/ru/tehnologii/vygodnaya-etika-kak-otvetstvennyj-ii-pomogaet-zarabatyvat-bolshe/, свободный (27.02.2025)
3) Bujold A., Roberge-Maltais I., Parent-Rocheleau X. et al. Responsible artificial intelligence in human resources management: a review of the empirical literature. AI Ethics 4. 2024 pp.1185–1200 1
4) Russel S., Norvig P. Artificial Intelligence. A modern approach. Pearson Education Limited. 2022 117 p.

5) The state of AI in early 2024: Gen AI adoption spikes and starts to generate value// McKinsey. 2024 [Электронный ресурс]. -Режим доступа: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/, свободный (10.02.2025)