

## СОЦИАЛЬНЫЕ ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Микулин К.А., студент*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Шепетюк В.В. – канд. фил. наук, доцент*

Статья посвящена социальным вызовам, обусловленным последствиями внедрения ИИ, начиная от возможной потери людьми рабочих мест, заканчивая отсутствием доступа к важнейшим ресурсам для жизни в связи с цифровым неравенством и манипуляциями сознанием. Рассматриваются пути минимизации рисков и угроз, альтернативные сценарии и направления развития технологических инноваций.

Искусственный интеллект стал неотъемлемой частью нашей жизни. Алгоритмы рекомендуют музыку, ставят медицинские диагнозы, одобряют кредиты и помогают судьям выносить решения. Однако, как отмечает редактор Financial Times Мадумита Мурджия, «слишком часто искусственный интеллект (далее – ИИ) приводит к эксплуатации, дискриминации и неравенству, особенно среди уязвимых групп населения» [1]. Рассмотрим основные социальные проблемы, связанные с внедрением ИИ.

Под угрозой автоматизации и замещения людей алгоритмами в настоящее время находятся около половины существующих профессий в развитых странах. Проблема не в полном исчезновении рабочих мест, а в их глубокой трансформации. Люди старшего возраста, жители отдаленных регионов, те, у кого нет доступа к качественному образованию, рискуют оказаться на обочине и стать маргинальными слоями населения. Авторы монографии «Искусственный интеллект в социальной сфере» подчеркивают: «Для ИИ изначально нехарактерны присущие человеческому разуму нелинейные решения, итерационные методы рассуждений. Следует признать объективные ограничения его применения в социально-гуманитарной сфере» [2].

Казалось бы, компьютер беспристрастен. Но алгоритмы обучаются на данных, содержащих человеческий опыт и стереотипы. В медицинской практике это особенно опасно: «Если в обучающих данных исторически недооценивались симптомы или чаще игнорировались потребности пациентов определенных этнических групп, LLM может перенять и воспроизвести эти нежелательные паттерны. Аналогично, гендерное смещение: долгое время медицинские исследования акцентировали внимание на мужском организме как "стандартом", игнорируя специфику женского» [3]. Относительно данной группы вопросов Эндрю Кокс и Мария Федотова добавляют, что ИИ порождает «множество этических проблем, касающихся пристрастности, прозрачности, объяснимости и ответственности, конфиденциальности, защиты и безопасности» [4].

Мы ежедневно сообщаем алгоритмам новые данные: геолокацию, поисковые запросы, историю покупок. ИИ анализирует указанные массивы, создавая наше цифровое досье, что дает возможность предугадывать наше поведение, реакции, потребности, перспективы дальнейшего выбора. На основании этих сведений в сочетании с другими факторами страховые компании, например, могут повышать цену полиса, если вы слишком часто заказываете фастфуд, а работодатели проявляют интерес к мониторингу активности и интересам сотрудников в нерабочее время. Анализ массивов данных на государственном уровне дает основания для создания целых систем социального рейтинга и управления общественными процессами, что уже успешно реализуется в Китайской Народной Республике. Таким образом, противоречивая грань между заботой и тотальным контролем постепенно стирается, последствия чего еще предстоит исследовать в будущем.

Особую угрозу представляют технологии deepfake и целый спектр связанных с ними социально-правовых проблем. Как отмечается в исследовании «Социально-правовая обусловленность формирования механизма уголовно-правового противодействия преступлениям, совершенным с использованием технологии deepfake», «дипфейк представляет собой сфальсифицированный аудио- и видеоконтент, сгенерированный с использованием генеративно-состязательных нейросетей» [5]. Использование таких технологий «способно причинить вред как правам и законным интересам отдельной личности, так и создавать более серьезные угрозы в информационном пространстве, затрагивая интересы всего государства» [5].

Е.В. Завгороднева и А.Г. Голощапова обращают внимание на проблему порнографических дипфейков: «высокая степень выбора женщин и несовершеннолетних в качестве жертв при создании глубоких порнодипфейков, а также степень морального и материального вреда, причиняемого вследствие генерирования и распространения сексуального интернет-насилия, основанного на изображениях требует принятия оперативных мер сдерживания потенциальных преступников выработки надлежащей общественной реакции на происходящее» [6].

ИИ требует огромных ресурсов, сосредоточенных, как правило, в богатых странах и размещенных там корпорациях. Бедные слои населения не участвуют в создании технологий, но всегда страдают от побочных эффектов их применения, когда весь мир перестраивается, модернизируется, а в каком-то месте отсутствуют возможности для следования новейшим трендам, стандартам и технологическим решениям. Таким образом, технологическое неравенство наслаивается на экономическое, порождая новые формы дискриминации различных групп населения, провоцируя ограничения для реализации их потребностей, интересов и даже порождая новые скрытые формы эксплуатации.

Современные нейросети устроены порой так сложно, а массивы обрабатываемых данных так огромны, что даже их создатели не всегда способны понять и реконструировать логику решений ИИ. Такое явление порождает эффект потенциальной непроверяемости выводов ИИ, что ведет к угрозе принятия его подсказок, советов или готовых решений в качестве идеальных. При этом даже в случае ошибки все еще не понятно, кто виноват: разработчик, пользователь или сам алгоритм. Юридическая система пока все еще не дает ответов на эти вопросы, актуальность и значимость которых нарастает.

Еще одной новой стороной проявления социальных рисков при использовании ИИ становится общение человека с чат-ботами, которое может вызывать эмоциональную привязанность, замещая реальные контакты. Анонимность, доступность, простота взаимодействия с системами ИИ при наличии у некоторых людей сложностей в самореализации в реальном мире по причине имеющихся страхов, ограничивающих убеждений или простой неуверенности в себе становятся очень привлекательными факторами. Алгоритмы соцсетей специально подбирают контент для удержания внимания, вызывая зависимость, интерес или тревогу. Особенно страдают от этого дети и подростки, жизненного опыта которых недостаточно для полноценного управления своим вниманием и поведением. Как следствие возникают эффекты социальной самоизоляции, социофобия, дезадаптация, социопатия, девиантное поведение и аномия.

Остановить технологический прогресс невозможно, однако, очевидно, что требуются принятие мер для управления складывающейся ситуацией в связи с внедрением ИИ в различных сферах общественной жизни. В первую очередь, на наш взгляд, необходимо внедрение решений, повышающих прозрачность алгоритмов ИИ. Под прозрачностью понимается способность объяснить, почему алгоритм принял то или иное решение, а также возможность для пользователей, регуляторов и независимых экспертов проверить, на каких данных строилась модель, какие признаки учитывались и нет ли в ней систематических искажений. Это означает внедрение механизмов объяснимости (explainability), обязательное документирование архитектуры и ограничений алгоритма, проведение регулярных аудитов и создание процедур оспаривания автоматических решений с участием человека. Во-вторых, огромное значение имеет законодательное регулирование всех сторон использования ИИ, что выражается в своевременном принятии нормативно-правовых актов, описывающих типичные ситуации его применения в повседневной жизни. В-третьих, важнейшую роль в обеспечении устойчивого будущего будет играть обучение людей навыкам применения естественного интеллекта в тех сегментах, где ИИ не может или не должен иметь решающее значение – речь о развитии и поддержании навыков критического мышления, этических стандартах поведения, навыках самоконтроля человека, ценностных ограничениях для разработчиков, а также о предвидении и управлении последствиями технологических решений. Как подчеркивают исследователи, «этическое регулирование ИИ требует расширенного комплексного подхода и междисциплинарного сотрудничества» [3].

В заключении отметим, что искусственный интеллект меняет не только технологии, но и все общество в целом – стиль жизни, стандарты, правила, ценности, ориентиры дальнейшего развития. Однако упрощение, оптимизация или трансформация жизни, не всегда означает ее улучшение. Актуальным остается главный вопрос – место и роль человека в новой технологической реальности, его способность сохранять свой статус автора осуществляющихся изменений и возможность нести ответственность за их последствия, направлять социальное развитие в требуемое русло. Проблемы, рождаемые внедрением ИИ невозможно игнорировать. Чтобы будущее было устойчивым, комфортным и безопасным для всех, обсуждать и решать эти проблемы нужно уже сейчас.

**Список использованных источников:**

1. Мурджия, М. Кодозависимые. Жизнь в тени искусственного интеллекта / М. Мурджия. - Москва: Синдбад, 2025. - 384 с.
2. Петрова, Т.Э. Искусственный интеллект в социальной сфере: монография / Т.Э. Петрова, Л.И. Селиванов, С.А. Соколовская [и др.]; под общ. ред. Т.Э. Петровой. – Москва: Русайнс, 2024. - 176 с.
3. Габидуллина, Л. Ф. Этика применения LLM-моделей в медицине и науке / Л. Ф. Габидуллина, М. Ю. Котловский // Медицинская этика. - 2025. - №3. - С. 9-13.
4. Кокс, Э. Исследования этических проблем применения искусственного интеллекта в библиотечном и информационном секторе / Э. Кокс, М.В. Федотова // Новости Международной федерации библиотечных ассоциаций и учреждений. - 2025. - №1(126). - С. 26-32.
5. Светанков, Д. Ю. Социально-правовая обусловленность формирования механизма уголовно-правового противодействия преступлениям, совершенным с использованием технологии deepfake / Д. Ю. Светанков // Российский следователь. - 2025. - № 6. - С. 47–51.
6. Завгороднева, Е.В. К вопросу о криминализации порнографических глубоких фейков / Е. В. Завгороднева, А.Г. Голощапова // Российско-азиатский правовой журнал. - 2025. - № 1. - С. 90–97.