УДК 338.24

КЛАССИФИКАЦИЯ РИСКОВ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

М. С. Абламейко[1)], С. В. Абламейко[2)]

[1)] Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, m.ablameyko@mail.ru
[2)] Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, ablameyko@bsu.by

Вопрос безопасного применения технологий ИИ становится все более важным для их безопасного применения. Это приводит к необходимости создания соответствующих условий, которые с одной стороны будут стимулировать их развитие, а с другой позволят ограничить их распространение в тех случаях, когда их применение способно будет нанести урон интересам личности, общества и государства. В этой связи, в последние годы все больше внимания уделяется анализу рисков применения систем ИИ. Связано это в первую очередь с автономностью их работы. Сбои работы таких систем могут привести к тяжелым последствиям. в данной работе предлагается классификация рисков систем ИИ.

Ключевые слова: искусственный интеллект; системы; риски; правовое регулирование.

Стремительное развитие информационных технологий привело к более широкому их применению во всех сферах жизни. В настоящее время технологии искусственного интеллекта (ИИ) способны принести широкий спектр экономических и социальных выгод как в отраслях экономики, так и в социальной жизни. Вместе с тем, вопрос безопасного применения технологий ИИ становится все более обсуждаемым в мировом сообществе, что в свою очередь приводит к необходимости создания соответствующих условий, которые с одной стороны будут стимулировать их развитие, а с другой позволят ограничить их распространение в тех случаях, когда их применение способно будет нанести урон интересам личности, общества и государства.

В этой связи в последние годы все больше внимания уделяется рискам применения систем ИИ. Связано это в первую очередь с автономностью работы систем ИИ. Сбои работы таких систем могут привести к тяжелым последствиям, в связи с чем обеспечение безопасности функционирования таких систем становится одной из первоочередных задач.

Основания для классификации систем искусственного интеллекта: 1) по степени автономности: автономные системы; встроенные системы; гибридные системы (обладают разной степенью независимости действий от участия человека и возможностей работать без вмешательства человека); 2) по степени автоматизации: автоматизированные системы; автоматические системы; 3) по архитектурному принципу: централизованные системы; распределенные системы; 4) по структуре и процессам обработки знаний: по модели знаний; по управлению знаниями; по методу обучения; 5) по специализации систем искусственного интеллекта: специализированные; универсальные; 6) по методам обработки информации: машинное обучение; глубокое обучение и др.; 7) по функциям управления: системы принятия (поддержки) решений; экспертно-аналитические системы; системы прогнозирования; 8) по степени опасности последствий; 9) по уровню конфиденциальности; 10) по видам деятельности; 11) по степени взаимодействия с человеком-оператором; 12) по степени адаптивности (возможность самообучения, позволяющая системе меняться во время использования). Возможно расширение видов классификации систем искусственного интеллекта. Возможно дополнение классификации как по новым основаниям, так и путем детализации классов по специализированным классификациям.

Классифицировать риски использования систем ИИ можно разными способами. По степени опасности риски систем ИИ классифицируются как: неприемлемый, высокий, ограни-

ченный и минимальный [1]. По степени нанесения вреда субъекту риски систем ИИ классифицируются как: риск причинения вреда человеку, организации, обществу и государству. В зависимости от сферы применения: в управлении, в финансовой сфере, в здравоохранении и системе медицинского обеспечения, в юриспруденции, в сфере образования и др.

В данной работе мы предлагаем классификацию рисков по важнейшим направлениям их использования. Исходя из этого, риски можно классифицировать следующим образом:

1. Социальные и этические риски

- Манипулирование общественным мнением (распространение пропаганды, использование технологии дип-фейков, дезинформация);
- Влияние на рынок труда (автоматизация рабочих мест, оптимизация количества работников и сокращение рабочих мест, необходимость перепрофилирования и обучения работников);
- Социальное неравенство (цифровая безграмотность, обеспечение доступа к технологиям, возрастной барьер);
- Вмешательство в частную жизнь (цифровая слежка, сбор и обработка персональных данных без согласия человека, профилирование);
- Дискриминация (социальный рейтинг);
- Общение с ИИ (тест Тьюринга);
- Зависимость от технологии.

2. Технологические риски

- Бесперебойная работа цифровой экосистемы государства с выявлением возможных внешних и внутренних угроз;
- Выход системы из-под контроля человека;
- Безопасность и приватность данных;
- Интерпретируемость результатов и предвзятость систем ИИ;
- Конфликт интересов (несовпадение целей ИИ и человека в случае постановки некорректной задачи);
- Использование иностранных платформ.

3. Военные риски

1. Автономные системы вооружений;
2. Применение ИИ в кибернетических и информационных операциях;
3. Военные «системы поддержки процесса принятия решений», основанные на ИИ.

Учитывая, что количество рисков применения технологий ИИ постоянно растет, считаем целесообразным подходить к решению данного вопроса комплексно. Безопасное использование технологий ИИ возможно при соблюдении определенных правил и стандартов. С технической точки зрения следует исходить из классификации систем ИИ и учитывать риск-ориентированный подход исходя из сферы применения. Необходимо разрабатывать, постоянно совершенствовать и применять технические стандарты и сертификацию систем ИИ. Также следует необходимо проводить испытания ИИ-систем и оценивать их результаты. Особое внимание следует уделять данным на которых система ИИ обучается.

С правовой точки зрения при принятии нормативных актов, принимаемых в сфере ИИ, должен выстраиваться баланс между инновациями и регулированием, чтобы развитие ИИ не имело негативных последствий от неконтролируемого применения. Полагаем, что на уровне подготовки актов законодательства в сфере ИИ следует фокусироваться на риск-ориентированном подходе, что позволит внести правовую определенность для различных сфер применения ИИ при определении требований для их использования, включая технические НПА, и минимизировать риски для человека, общества и государства. Наряду с принятием законодательных актов, ограничивающих применение ИИ, следует учитывать этические нормы, т.е. применять «мягкое право».

Также следует отметить, что регулирование данной сферы должно не только основываться на международных правилах, но и учитывать национальную специфику, что особенно касается военной сферы.

Таким образом, для преодоления рисков следует уделять внимание: обеспечению кибербезопасности, принятию законодательных актов и этических стандартов, обеспечению социальной поддержки и обучению пользования технологиями ИИ населения, развитие прозрачности в принятии решений ИИ.

Определение четких критериев отнесения системы ИИ к той или иной категории риска предоставит возможность классифицировать системы ИИ и принимать отраслевые стандарты, исходя из общих требований. В качестве критериев оценки рисков, исходя из международного опыта, можно определить следующие: оценка цели и сферы применения системы ИИ; степень автономности системы и механизмы контроля со стороны человека; сложность системы, включая прозрачность и объяснимость; технология, лежащая в основе системы и др.

Библиографические ссылки

1. Абламейко М.С. К вопросу о разработке Кодекса этики искусственного интеллекта. // Динамика правоустановления и правореализации в сфере публично-правовых отношений: сборник научных статей / Национальный центр законодательства и правовых исследований Республики Беларусь; О.И.Чуприс (гл.ред.) [и др.]. Минск: Колорград, 2023. Вып. 5. С.231-237.