



DOI: <https://doi.org/10.15688/ek.jvolsu.2023.4.1>

UDC 334.78
LBC 65.291.2

Submitted: 05.10.2023
Accepted: 26.10.2023

THE PHENOMENON OF RISK UNDER THE APPLICATION OF ARTIFICIAL INTELLIGENCE ALGORITHMS

Roman M. Kachalov

Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russian Federation

Yulia A. Sleptsova

Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russian Federation

Abstract. With the advent of large language models, researchers became interested in artificial intelligence algorithms and their application in the activities of enterprises. Such algorithms have already made a significant contribution in the fields of medicine, transport and finance. It is now necessary to understand what side effects artificial intelligence algorithms may cause once they are integrated into the production of goods and services. The purpose of this work is to identify risk factors for the use of algorithms and artificial intelligence systems, assess their impact on the level of risk and offer tools to reduce the level of risk. As a result of the conducted research, the term “artificial intelligence” was clarified; algorithms for analytical, functional, interactive, textual and visual artificial intelligence were identified. The modern artificial intelligence system represents the next wave of enterprise automation. Using the methods of system economic theory and operational theory, the risk factors for the use of artificial intelligence algorithms in the object, process, design and environmental subsystems of the enterprise are classified. Anti-risk management actions are proposed to reduce the likelihood of a risk situation or to compensate for the damage caused. The changes in the activities of large companies in the field of information technology related to the change in the concept of copyright and intellectual property are described. The limitations of the introduction and application of artificial intelligence algorithms associated with the inertia of managers’ thinking and the complexity of restructuring business processes at the enterprise are shown. The necessity of regulation in the field of the introduction of artificial intelligence systems, the definition of clear boundaries and restrictions to ensure the safety of individuals and society as a whole, and their compliance with ethical standards are substantiated.

Key words: risk factors, level of risk, artificial intelligence, large language models, system economic theory, object subsystem, process subsystem, design subsystem, environmental subsystem.

Citation. Kachalov R.M., Sleptsova Yu.A. The Phenomenon of Risk Under the Application of Artificial Intelligence Algorithms. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Ekonomika* [Journal of Volgograd State University. Economics], 2023, vol. 25, no. 4, pp. 5-16. (in Russian). DOI: <https://doi.org/10.15688/ek.jvolsu.2023.4.1>

УДК 334.78
ББК 65.291.2

Дата поступления статьи: 05.10.2023
Дата принятия статьи: 26.10.2023

ФЕНОМЕН РИСКА В УСЛОВИЯХ ПРИМЕНЕНИЯ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Роман Михайлович Качалов

Центральный экономико-математический институт РАН, г. Москва, Российская Федерация

Юлия Анатольевна Слепцова

Центральный экономико-математический институт РАН, г. Москва, Российская Федерация

Аннотация. С появлением больших языковых моделей у исследователей возрос интерес к алгоритмам искусственного интеллекта и их применению в деятельности предприятий. Такие алгоритмы уже внесли свой значительный вклад в сферах медицины, транспорта и финансов. В настоящее время необходимо понять, какие побочные эффекты алгоритмы искусственного интеллекта могут вызвать после их интеграции в производство продуктов и услуг. Целью данной работы является выявление факторов риска применения таких алгоритмов и систем, оценка их влияния на уровень риска и предложение инструментов воздействия для снижения уровня риска. В результате проведенного исследования был уточнен термин «искусственный интеллект», выделены алгоритмы аналитического, функционального, интерактивного, текстового и визуального искусственного интеллекта, составляющие современную систему искусственного интеллекта, которая представляет собой следующую волну автоматизации деятельности предприятий. С использованием методов системной экономической теории и операциональной теории классифицированы факторы риска применения алгоритмов искусственного интеллекта в объектной, процессной, проектной и средовой подсистемах предприятия и предложены антирисковые управленческие воздействия для уменьшения вероятности возникновения ситуации риска или для компенсации причиненного ущерба. Описаны изменения в деятельности крупных компаний в области информационных технологий, связанные с изменением понятия авторского права, интеллектуальной собственности в настоящее время. Показаны ограничения внедрения и применения алгоритмов искусственного интеллекта, связанных с инерционностью мышления руководителей и сложностью перестройки бизнес-процессов на предприятии. Обоснована необходимость регулирования в сфере внедрения систем искусственного интеллекта, определения четких границ и ограничений для алгоритмов искусственного интеллекта, для обеспечения безопасности отдельных людей и общества в целом, их соответствие этическим нормам.

Ключевые слова: факторы риска, уровень риска, искусственный интеллект, большие языковые модели, системная экономическая теория, объектная подсистема, процессная подсистема, проектная подсистема, средовая подсистема.

Цитирование. Качалов Р. М., Слепцова Ю. А. Феномен риска в условиях применения алгоритмов искусственного интеллекта // Вестник Волгоградского государственного университета. Экономика. – 2023. – Т. 25, № 4. – С. 5–16. – DOI: <https://doi.org/10.15688/ek.jvolsu.2023.4.1>

Введение

Активные сторонники «искусственного интеллекта» (ИИ), в отличие от человеческого – естественного интеллекта, полагают, что такие технологии предлагают обществу и предприятиям дополнительные ценности и новые формы добавленной стоимости. Надо заметить, что сами по себе технологии не определяют будущую деятельность предприятий и их сотрудников, клиентов, представителей поставщиков. Технологии являются своеобразными инструментами, при помощи которых создаются продукты, которыми начинают пользоваться люди, и продукты, ставшие необходимыми, действительно могут поменять используемые бизнес-процессы и деловую среду.

Можно сказать, что так и случилось с алгоритмами создания и обучения больших языковых моделей – Large Language Models (LLM), продукт на основе которых очень быс-

тро стал популярным в течение последнего года (речь идет о продукте ChatGPT от компании OpenAI). Пока большинство новых функций этих моделей находится на различных этапах тестирования, и доступ к ним ограничен, но благодаря отчетам о ходе испытаний можно сделать выводы об изменении привычной рабочей среды сотрудников в офисах компаний [“So What If ChatGPT Wrote It?” ... , 2023]. В Китае у ведущих компаний, таких как Tencent, Alibaba, Baidu и Huawei, также существуют свои LLM; эти модели применяются предприятиями, использующими в своей деятельности китайский язык. Таким образом, LLM не только обрабатывают и создают тексты на языках народов мира, но и формируют новые инструменты для работы дизайнеров и программистов.

Применение новых алгоритмов может быть связано с существенным увеличением уровня риска. В первую очередь с такими

факторами риска могут столкнуться компании, которые разрабатывают и внедряют системы на основе ИИ. На следующем этапе рискам могут быть подвержены пользователи и иные заинтересованные стороны, которые вовлечены во взаимодействие с продуктами на основе ИИ. Если пользователям причинен какой-либо ущерб и они имеют достаточную власть, негативная обратная связь с их стороны может повлиять на компании-разработчики, создающие продукты с использованием ИИ, особенно на те, которые имеют прямые контакты с пользователями.

Можно выделить три категории таких систем ИИ в зависимости от уровня риска, который несет за собой их использование:

- неприемлемый уровень риска;
- высокий уровень риска;
- средний или минимальный уровень риска.

Уровень риска может считаться неприемлемым в случае использования систем, включающих алгоритмы ИИ, если они применяются для классификации людей на слои по «чувствительным признакам»: пол, раса, гражданство, религия и т. п. Применение таких систем должно быть запрещено.

Высоким уровнем риска обладают алгоритмы ИИ и платформы, используемые в критической инфраструктуре; системы, которые могут быть использованы в образовательных целях и оценивающие учащихся или определяющие, кому предоставить доступ к обучению; алгоритмы оценки работников как при приеме на работу, так и уже сотрудничающих с организацией [Лобачева и др., 2021]; системы оценки кредитоспособности человека.

Таким образом, для снижения уровня риска желательно разработать этические нормы не только на этапе создания программного обеспечения, но и на этапе использования искусственных интеллектуальных систем [Качалов и др., 2023].

Целью настоящего исследования является идентификация факторов риска применения алгоритмов искусственного интеллекта, выявление их влияния на уровень риска и предложение инструментов управленческого воздействия, направленных на снижение уровня риска в деятельности предприятия.

Для достижения указанной цели необходимо решить следующие задачи:

- сформулировать понятие «искусственный интеллект»;
- выявить факторы риска в деятельности предприятия и предложить антирисковые управленческие воздействия при использовании данных технологий с учетом их потенциального влияния на все сферы функционирования предприятия.

Алгоритмы искусственного интеллекта: определение, виды

Само понятие «искусственный интеллект» трактуется по-разному, и поэтому его точные контуры очертить довольно непросто. Актуальность этого термина обсуждается и регулярно ставится под сомнение.

С одной стороны, под ИИ понимается некоторая область науки и техники, которая занимается пониманием рационального поведения и созданием искусственных систем, воспроизводящих это поведение [Shapiro, 1992]. С другой стороны, ИИ можно определить, как область науки, изучающую интеллектуальное поведение в вычислительных терминах [Levesque, 2014]. Для некоторых исследователей в этой области долгосрочной целью создания ИИ является своеобразная автоматизация интеллекта человеческого уровня [Nilsson, 2005]. Предполагалось, в частности, что все формы интеллекта (рассуждение, понимание языка коммуникаций между людьми, поддержка принятия решений и т. д.) состоят из манипулирования символами в соответствии с формальными правилами логического вывода, а, с другой стороны, что поведение определяется некоторыми мысленными представлениями [Collins, 2018]. Некоторые признанные специалисты в этой области даже считают, что на настоящем этапе развития технологий системы искусственного интеллекта характеризуются их возможностями распознавания образов, текстов, речи и т. п. и способностями таких систем учиться на основе опыта, приспосабливаться к новым входным данным и выполнять недостаточно ясно сформулированные задачи [Le Cun, 2019; Julia, 2019]. Могут быть выделены различные виды алгоритмов ИИ: аналитический, функциональный, интерактивный,

текстовый и визуальный ИИ [Sarker, 2022], которые рассматриваются далее.

Алгоритмы *аналитического ИИ* используются в процессе выявления трендов, их интерпретации и осмысленного обобщения на основе массивов данных. Таким образом, аналитический ИИ стремится обнаружить новые закономерности, отношения или зависимости в данных, а также помочь в принятии решений на основе данных. Таким образом, в сфере современной бизнес-аналитики алгоритмы ИИ могут предоставить информацию и рекомендации руководству предприятия с помощью аналитической обработки. Различные методы машинного обучения могут использоваться для построения аналитических моделей ИИ для решения конкретных проблем на предприятии. Например, для оценки факторов бизнес-риска, можно использовать аналитические модели, основанные на дата-сетах [Sarker, 2021].

Алгоритмы *функционального ИИ* работают аналогично алгоритмам аналитического ИИ, в части исследования больших объемов данных и идентификации закономерностей, но, с другой стороны, такие алгоритмы участвуют в выполнении действий, а не в выдаче рекомендаций. Например, эти алгоритмы применяются в робототехнике и приложениях, связанных с Интернетом вещей (IoT) для принятия немедленных действий [Иванов и др., 2018].

Алгоритмы *интерактивного ИИ* используются предприятиями для автоматизации интерактивного общения в коммерческой сфере или для поддержки принятия управленческих решений. Например, модель интерактивного ИИ может быть полезна для запуска чат-бота для общения с клиентами, или использования персонального помощника [A Systematic Literature ... , 2023].

Алгоритмы *текстового ИИ* уже обсуждались выше, когда речь шла о больших языковых моделях, то есть такие алгоритмы обычно применяются для обработки естественного языка, с помощью технологий распознавания текста, преобразования речи в текст, машинного перевода или генерации контента. Предприятие, например, может использовать LLM для поддержки корпоративного (внутреннего) хранилища знаний, предоставления соответ-

ствующих услуг или для реагирования на запросы потребителей [Sallam, 2023].

Алгоритмы *визуального ИИ* используются для распознавания, классификации и сортировки элементов, а также для конвертации изображения и видео. Эти алгоритмы применяются в таких технологиях, как компьютерное зрение и дополненная реальность, или предприятиями в сфере безопасности и в области образования [Chiu, 2023].

При оформлении регуляторных норм и правил потребовалось более строгое введение понятия *системы искусственного интеллекта* как некоторого программного обеспечения, развивающегося на основе применения технологий машинного обучения, статистических и иных подходов и методов, которое может производить тексты, рекомендации, прогнозы или решения.

При расширении такого толкования термина «искусственный интеллект» можно определить *систему ИИ*, как основанную на машинном обучении систему, функционирующую на различных уровнях автономии, имеющую способность генерировать предиктивную аналитику и рекомендации, которые могут оказывать физическое или виртуальное влияние на решения, принимаемые людьми, как в частной, так и деловой сферах.

Главной особенностью методов машинного обучения является построение системы рассуждений непосредственно из большого массива данных, без явных правил для генерации результата процесса [Лисовский, 2020]. Универсальность таких методов делает их очень привлекательными для широкого спектра применений.

Кроме того, разработчики технологий машинного обучения с самого начала приняли открытый подход к сотрудничеству и распространению большого набора ресурсов, от программного обеспечения до наборов данных, документации, свободно доступных каждому. Этот подход способствовал росту популярности методов машинного обучения в инженерных и бизнес-сообществах и использованию преимуществ огромных объемов данных, собранных в цифровых системах.

С другой стороны, у пользователей возникли конкретные этические требования предоставлять понятные объяснения при приня-

тии решений без участия человека. Проблема применения этических норм при автоматизации процессов принятия решений включает в себя два основных класса проблем. Во-первых, это создание формализованных представлений таких норм, а во-вторых, выбор соответствующего математического аппарата для работы с этими нормами, то есть задачи сопоставления, измерения, анализа и т. п. [Карпов и др., 2018].

Особенностью применения алгоритмов искусственного интеллекта является их сервисный характер, что приводит к необходимости косвенного оценивания их эффективности через терминальный положительный эффект [Юсупов и др., 2017] на предпринимательскую деятельность, в интересах которой данная система была создана или адаптирована.

В качестве оценки эффективности системы искусственного интеллекта по критерию пригодности может быть использована вероятность превышения достигнутым положительным эффектом минимального порогового значения, которое можно трактовать как вероятность выполнения стоящей перед системой искусственного интеллекта задачи. Поэтому дополнительная величина к такому пороговому значению может представлять собой вероятность невыполнения поставленной задачи, что совпадает с определением понятия уровня риска.

Феномен риска применения алгоритмов ИИ в подсистемах предприятия

Очередная волна компьютерной автоматизации профессиональных задач представляет собой одно из наиболее обсуждаемых последствий применения алгоритмов ИИ, особенно потому, что этот этап затрагивает все больше и больше предприятий и отраслей (в секторе услуг, промышленности, сельском хозяйстве, образовании, финансах, медицине и т. д.). Современный период компьютеризации в основном характеризуется ее распространением на когнитивные аспекты человеческой деятельности, такие как принятие решений и коммуникации. Такое расширение автоматизации с помощью алгоритмов ИИ стало возможным благодаря распространению использования

специализированных инструментов для сбора данных, необходимых для обучения систем ИИ в рабочих ситуациях деятельности экономической системы [Jarrahi et al., 2021].

Представление о четырехэлементной типологии экономических систем опирается на особенности локализации систем в пространстве и времени [Клейнер, 2022]. Согласно теории Г.Б. Клейнера в экономике по пространственно-временному признаку можно выделить четыре типа социально-экономических систем: объектные, средовые, проектные, процессные. В границах предприятия, в соответствии с тем же пространственно-временным подходом, можно выделить аналогичные четыре подсистемы объектного, средового, проектного и процессного типов [Качалов и др., 2022]. Основываясь на этом утверждении, можно выявлять факторы риска и разрабатывать соответствующие антирисковые управленческие воздействия в пределах каждой из этих подсистем относительно автономно.

В подсистему объектного типа включаются материальные и нематериальные активы – оборудование, готовая продукция, сырье, материалы, денежные средства, используемые патенты и т. п. Отдельно может быть рассмотрен правовой статус используемых на предприятии активов (в том числе оборудования и основных средств, полученных по лизинговым соглашениям), а также имущества на правах собственности.

Фактором риска использования гибридных продуктов предприятия, то есть сочетающих в себе как инновационные достижения в технике, так и встроенные программные комплексы, могут стать ошибки при разработке и опытной эксплуатации изделий, допущенные разными специалистами. Так, например, внедрение новейших медицинских имплантатов может служить хорошим примером многосторонней оценки и управления рисками ответственного применения алгоритмов ИИ.

Результаты оценки зависят от процессов управления уровнем риска, проводимых с точки зрения каждой заинтересованной стороны. Заинтересованные стороны включают в себя производителей медицинской продукции, проводящих тщательно спланированные пилотные исследования, при активном участии медицинских работников, пациентов и организа-

ций по защите прав пациентов [Clarke, 2019]. Таким образом, обобщая, можно сказать, что понятие «заинтересованные стороны» шире, чем просто пользователи систем ИИ. В него входят не только клиенты, эксплуатирующие продукцию со встроенными системами ИИ, но также и «любые другие лица, группы или предприятия, действия которых могут повлиять на развитие и использование системы прямо или косвенно» [Pouloudi et al., 1997].

С точки зрения владения нематериальными активами у компаний в области создания и обучения систем ИИ за последние годы произошли серьезные сдвиги. Так называемый открытый код (open source) фактически изменил представление о промышленных секретах. Еще некоторое время назад информационные системы, созданные на предприятиях, были самой важной коммерческой тайной. Разработки компании IBM или компании Microsoft на стадии зарождения операционной системы Windows становились интеллектуальной собственностью, то есть таким нематериальным активом, который невозможно воспроизвести. Оценка стоимости такого актива влияла на стоимость самой компании или ее ценных бумаг.

В настоящее время компании и предприятия начинают выкладывать свои новейшие системы в общий доступ, так как для стратегического развития предприятия самым важным ресурсом становится талант сотрудников, которые в нем работают [Кабалина и др., 2018]. Предприятия вынуждены конкурировать за лучших программистов с крупнейшими производителями информационных систем, а такие специалисты прежде чем дадут согласие на трудовой договор, хотят ознакомиться с продуктом, вынуждая работодателей вслед за компанией Google или компанией Яндекс выкладывать свои системы в открытый доступ. Таким образом алгоритмы машинного обучения или системы ИИ становятся публично доступны. Фактором риска в таком случае может стать переобучение алгоритма под контролем специалиста с преступным умыслом, например, с помощью таких инструментов, как универсальный алгоритм Tensor Flow, разработанный в компании Google. Другая модель Cat Boost, принадлежащая российской компании Яндекс, тоже имеет открытый дос-

туп. Теоретически обществу нужно контролировать людей, потому что, когда речь заходит об интеллектуальном мире, машинах, алгоритмах, информации, у потенциального злоумышленника есть возможность использовать наработки, содержащиеся в открытом доступе для обучения алгоритмов ИИ чему-либо противозаконному.

Подсистема процессного типа состоит из технологических и бизнес-процессов, действующих для обеспечения функционирования предприятий, которые характеризуются устойчивостью и многократным повторением.

Области применения алгоритмов ИИ в подсистемах процессного типа предприятий продолжают расширяться при улучшении взаимоотношения с клиентами, логистика, прогнозный анализ, диагностика, анализ больших баз данных и т. д. В этом контексте ИИ чаще всего рассматривается как набор технологий, которые могут принести многочисленные преимущества, особенно с точки зрения производительности (оптимизация внутренних процессов, скорость выполнения задач, повышение производительности и т. д.), а иногда и с точки зрения облегчения работы или даже снижения трудоемкости за счет автоматизации утомительных или повторяющихся задач.

Одним из факторов риска при использовании систем ИИ в процессной подсистеме предприятия может быть непрозрачность как для представителя самого предприятия, так и для контрагента, то есть какая конкретно часть процесса (коммуникации или действия) выполняется ИИ. Новые перспективы, открываемые системами ИИ, требуют переосмысления разделения труда между людьми и алгоритмами, поскольку прогресс в этой области означает, что системы ИИ становятся более эффективными, чем традиционные сотрудники, во многих областях, таких как диагностика или прогнозный анализ. Например, нам регулярно говорят, что системы искусственного интеллекта лучше врачей выявляют или оценивают вероятность появления заболеваний.

Фактором риска интегрирования систем ИИ в процессы, используемые в чувствительных областях, таких как безопасность жизнедеятельности людей, постановка медицинских диагнозов и т. д., может стать распределение

ответственности за ошибку между разработчиками такого программного комплекса и специалистами, обучающими такие системы. Иными словами, в случае серьезной ошибки, повлекшей за собой существенный ущерб здоровью или жизни людей, сложно будет возложить юридическую ответственность за это на программный комплекс, а не на реального человека.

В программных комплексах с использованием ИИ пользователям должна быть предоставлена возможность обходить или игнорировать системные решения, когда это необходимо (например, в случае необычных ситуаций или новых условий, которые не ожидаются или не могут управляться алгоритмами). Так как при резких изменениях в методах ведения бизнеса, например, связанных с пандемией нового заболевания, алгоритмы, обученные на массивах «исторических» данных, показывающих как предприятие или организация работали до начала пандемии, перестали быть релевантными текущей ситуации [Качалов и др., 2022а].

В качестве антирискового управленческого воздействия должно применяться положение, что ответственность за решение, принятое с помощью систем ИИ за действие или за процесс коммуникации, должна быть возложена на конкретного компетентного специалиста, оценивающего такое решение [The Ten Commandments ... , 2021].

Таким образом, контроль и мониторинг подготовки управленческих решений, действий или процессов коммуникаций, совершенных с помощью алгоритмов ИИ, не должны выполняться программными комплексами, а только человеком.

Подсистемы проектного типа образуют совокупность реализуемых в рамках предприятия проектов, включая разработку продуктов или услуг, реорганизацию информационных систем и т. п.

Использование систем ИИ, генерирующих проектные решения, существенно трансформирует проектную деятельность и требует от специалистов, например, в области разработки программных продуктов или дизайна новых навыков. С помощью этих инструментов деятельность такого творческого специалиста больше не является проектной деятель-

ностью в «классическом» смысле (создание эскиза, его доработка и модификация и т. д.), а все больше структурируется с помощью следующих действий [Autonomous Tools ... , 2018]:

- определение параметров и ограничений, которым должна следовать система (формирование запроса);

- оценка и тестирование решения, сгенерированного алгоритмами ИИ;

- корректировка параметров или алгоритмов в соответствии с оценкой.

Таким образом, применение систем ИИ в проектной подсистеме предприятия предполагает хорошее понимание руководством и сотрудниками функционирования этих систем, и, при необходимости, изменение предлагаемых ими решений.

Одним из *факторов риска проектной системы* предприятия будет фактор кадрового риска, в случае, если имеющиеся в распоряжении предприятия специалисты не овладеют навыками работы с новыми технологиями.

Антирисковыми управленческими воздействиями для минимизации последствий реализации такого фактора риска могут быть стимулирование обучения сотрудников работе с новыми цифровыми технологиями и/или создание кадрового резерва из числа работников предприятия, возможно не связанных с работой креативных подразделений, но заинтересованных в обучении использованию алгоритмов ИИ.

Подсистемы среднего типа представляют собой сложившиеся в пределах предприятия культурные традиции, социальные условия и другие неформальные институты, а также действующее федеральное и местное законодательство, регулирующее деятельность по производству, реализации и безопасности продукции и услуг для потребителей и других членов общества.

При запуске беспилотного транспорта руководство предприятия, применяющего алгоритмы ИИ, должно учитывать, что беспилотные автомобили влияют не только на пассажиров внутри такого автомобиля, но и на водителей и пассажиров других транспортных средств и пешеходов.

Фактором риска в этом случае будет возможность причинения вреда участникам

дорожного движения, находящимся поблизости от беспилотного автомобиля.

При использовании алгоритмов ИИ в потребительском кредитовании, медицинском страховании или социальном обеспечении существует вероятность возникновения факторов риска дискриминации клиентов по различным признакам. Существуют опасения, что критерии выбора процентной ставки, того или иного плана страхования, или суммы социаль-

ного пособия могут быть недостаточно прозрачными.

В случае применения алгоритмов ИИ иногда сложно классифицировать определенно факторы риска конкретной подсистемы, так как существует сильная взаимосвязь и взаимовлияние подсистем предприятия, причем такая связь осуществляется именно с помощью новых цифровых технологий.

Таблица. Примеры факторов риска применения алгоритмов ИИ в подсистемах предприятия

Table. Examples of risk factors for the use of AI algorithms in enterprise subsystems

Фактор риска	Антирисковое управленческое воздействие
Объектная подсистема	
Фактор риска неточности текста, диагноза библиографической ссылки при использовании LLM-модели	– Создание и/или использование специализированных, а не общих моделей. – Сертификация специализированных LLM-моделей со стороны профессионального сообщества
Фактор риска отказа запуска программного комплекса, созданного с помощью алгоритма ИИ	Использование высококлассного разработчика программных комплексов
Фактор риска низкой производительности программного продукта, разработанного с использованием системы ИИ	
Процессная подсистема	
Фактор риска переноса данных, вводимых сотрудниками предприятия, в большие языковые модели и раскрытия коммерческих секретов предприятия другим пользователям	– Гарантии создателя модели «ChatGPT Enterprise» компании OpenAI. – Предложения компании Microsoft по защите авторских прав для покрытия судебных издержек в случае, если компании будет предъявлен иск за нарушение авторских прав в результате использования продуктов Microsoft Copilot AI
Фактор риска некорректного обучения моделей LBM (Large Behavior Models)	Дополнительное тестирование и набор больших объемов наблюдений за выполнением человеком той или иной физической задачи в реальном мире, а затем с использованием механизма обратной связи программировать модель на гибкое выполнение этой задачи
Проектная подсистема	
Фактор риска возникновения ошибок моделей (LLM) с мультимодальным вводом при генерации обобщений, связанных с изображениями, и реакции модели на уточнения в графическом виде (например, дорисованные человеком «от руки» поверх фотографии, графика или чертежа)	Дополнительное тестирование моделей и контроль полученных результатов
Конкретный набор навыков, который есть у человека для коммуникации с моделью ИИ, будет устаревать по мере выхода новой версии модели	– Непрерывное или периодическое обучение сотрудников. – Адаптация новых моделей к интуитивно понятному для человека интерфейсу
Средовая подсистема	
Фактор риска присутствия «двойника» Duet AI от компании Google на онлайн-встрече	Проведение встреч в очном формате (хотя не всегда это возможно)
Фактор риска избытка информации, и, как следствие, замедление выполнения задач сотрудниками предприятия и использования приложения на основе ИИ для создания заметок, таких как Capacities или Google NotebookLM, как единственного источника информации для деятельности	Контроль настроек приложения и регулярное тестирование получающегося результата

Примечание. Составлено авторами.

Примеры факторов риска использования систем ИИ с предполагаемыми антирисковыми управленческими воздействиями приведены в таблице.

Выводы

1. Показано, что для эффективного применения систем ИИ потребуется психологически тяжелая для многих руководителей задача делегирования некоторых трудовых функций и полномочий алгоритмам искусственного интеллекта, заменяющим персонал. Учитывая недавнее появление и быстрое развитие технологии и алгоритмов ИИ, их интеграция в бизнес уже вызвала ряд проблем. Некоторые проблемы, связанные с нарушением конфиденциальности и авторских прав, решаются.

2. Установлено, что в настоящее время появился консенсус в том, что внедрение алгоритмов ИИ приведет к более или менее глубокой реконфигурации процессов и проектов в деятельности предприятий. Один из важных выводов прошлых исследований в области автоматизации заключается в том, что люди по-прежнему играют важную роль в рабочей среде, особенно для реагирования на непредвиденные события или при сбоях в работе систем ИИ. То есть для осуществления антирисковых управленческих воздействий потребуются не только алгоритмы, но и реальные сотрудники.

3. Показано, что применение систем ИИ не только сопряжено с риском, но такие алгоритмы могут быть также использованы при разработке антирисковых управленческих воздействий, в том числе для решения этических проблем. Например, для написания сложных деловых писем, затрагивающих деликатные темы, прежде чем проект, инициированный предприятием, сможет развиваться дальше, или отклоняющих чье-либо предложение, не оттолкнув его автора. Почтовое приложение Superhuman, созданное с использованием алгоритмов ИИ, предлагает функцию написания электронных писем и сообщений с учетом контекста предыдущей переписки.

4. Установлено, что, возможно, следующим шагом в развитии алгоритмов ИИ будет переход от LLM- или LBM-моделей к инте-

ративным системам ИИ. В отличие от генеративных моделей, которые в основном сосредоточиваются на создании текста, изображения или иного контента, интерактивные алгоритмы ИИ будут способны выполнять конкретные задачи по запросу пользователя, в том числе взаимодействуя с другими программами, сервисами и даже с сотрудниками для достижения поставленных целей. Такие системы ИИ могут привести серьезные изменения в технологический ландшафт предприятия, предоставляя ИИ права действовать автономно в рамках заданных параметров, то есть до некоторой степени «агентность». С одной стороны, такое развитие будет связано не только с технологическим прогрессом, но и откроет широкие социальные и экономические перспективы. А с другой стороны, необходимо строгое регулирование в новой сфере, должны быть четко определены границы и ограничения, которые не могут пересечь системы ИИ, чтобы обеспечить безопасность как отдельных людей, так и общества в целом, и соответствие этическим нормам.

5. Показано, что существует вероятность того, что массовое внедрение алгоритмов ИИ, несмотря на активное привлечение общественного внимания, потребует длительного времени, также как потребовалось много лет на предыдущую цифровую трансформацию бизнес-процессов. Выгоды получают те предприятия, которые смогут адаптироваться к применению новых технологий ИИ в кратчайшее время.

СПИСОК ЛИТЕРАТУРЫ

- Иванов, Д. А. Анализ тенденций изменения принципов управления предприятиями в условиях развития технологий индустрии 4.0 / Д. А. Иванов, М. А. Иванова, Б. В. Соколов // Тр. СПИИРАН. – 2018. – Вып. 5 (60). – С. 97–127. – DOI: 10.15622/sp.60.4
- Кабалина, В. И. Контекстуализация управления талантами в российских ИТ-компаниях / В. И. Кабалина, О. В. Мондрус // Российский журнал менеджмента. – 2018. – Т. 16, № 1. – С. 5–36. – DOI: <https://doi.org/10.21638/11701/spbu18.2018.101>
- Карпов, В. Э. К вопросу об этике и системах искусственного интеллекта / В. Э. Карпов, П. М. Готовцев, Г. В. Ройзензон // Философия и обще-

- ство. – 2018. – Т. 2 (87). – С. 84–105. – DOI: 10.30884/jfio/2018.02.07
- Качалов, Р. М. Бизнес-экосистемы в кризисных условиях: выявление факторов риска / Р. М. Качалов, Ю. А. Слепцова // *Российский журнал менеджмента*. – 2022. – Т. 20, № 2. – С. 155–171. – DOI: <https://doi.org/10.21638/spbu18.2022.201>
- Качалов, Р. М. Управление риском в деятельности бизнес-экосистем в период кризисов / Р. М. Качалов, Ю. А. Слепцова // *Вестник Волгоградского государственного университета. Экономика*. – 2022а. – Т. 24, № 2. – С. 146–159. – DOI: <https://doi.org/10.15688/ek.jvolsu.2022.2.12>
- Качалов, Р. М. Этические проблемы применения алгоритмов искусственного интеллекта в управлении предприятием / Р. М. Качалов, Ю. А. Слепцова // *Системный анализ в проектировании и управлении*. – 2023. – Т. 26 (2). – С. 16–23. – DOI: 10.18720/SPBPU/2/id23-77
- Клейнер, Г. Б. Информационная теория факторов управления экономическими организациями / Г. Б. Клейнер // *Российский журнал менеджмента*. – 2022. – Т. 20, № 4. – С. 461–481. – DOI: <https://doi.org/10.21638/spbu18.2022.401>
- Лисовский, А. Л. Применение нейросетевых технологий для разработки систем управления / А. Л. Лисовский // *Стратегические решения и риск-менеджмент*. – 2020. – № 11 (4). – С. 378–389. – DOI: 10.17747/2618-947X-2020-4-378-389
- Лобачева, А. С. Этика применения искусственного интеллекта в управлении персоналом / А. С. Лобачева, О. В. Соболев // *E-Management*. – 2021. – Т. 4 (1). – С. 20–28. – DOI: 10.26425/2658-3445-2021-4-1-20-28
- Юсупов, Р. М. Особенности оценивания эффективности информационных систем и технологий / Р. М. Юсупов, А. А. Мусаев // *Тр. СПИИРАН*. – 2017. – № 2 (51). – С. 5–34. – DOI: 10.15622/sp.51.1
- Algorithmic Management in a Work Context / М. Н. Jarrahi [et al.] // *Big Data & Society*. – 2021. – Vol. 8 (2), art. 20539517211020332. – DOI: 10.1177/20539517211020332
- A Systematic Literature Review of Information Security in Chatbots / J. Yang, Y. L. Chen, L. Y. Por, C. S. Ku // *Applied Sciences*. – 2023. – Vol. 13 (11). – P. 6355. – DOI: 10.3390/app13116355
- Autonomous Tools and Design: A Triple-Loop Approach to Human-Machine Learning / S. Seidel, N. Berente, A. Lindberg, K. Lyytinen, J. V. Nickerson // *Communications of the ACM*. – 2018. – Vol. 62, iss. 1. – P. 50–57. – DOI: 10.1145/3210753
- Chiu, T. K. F. The Impact of Generative AI (GenAI) on Practices, Policies and Research Direction in Education: A Case of ChatGPT and Midjourney / T. K. F. Chiu // *Interactive Learning Environments*. – 2023. – DOI: 10.1080/10494820.2023.2253861
- Clarke, R. Principles and Business Processes for Responsible AI / R. Clarke // *Computer Law & Security Review*. – 2019. – Vol. 35, iss. 4. – P. 410–422. – DOI: 10.1016/j.clsr.2019.04.007
- Collins, H. *Artificial Intelligence: Against Humanity's Surrender to Computers* / H. Collins. – Polity Press, 2018.
- Julia, L. *L'intelligence artificielle n'existe pas* / L. Julia. – Paris : Éditions First. 2019.
- Le Cun, Y. *Quand la machine apprend: la révolution des neurones artificiels et de l'apprentissage profond* / Y. Le Cun. – Odile Jacob, 2019.
- Levesque, H. J. On Our Best Behaviour / H. J. Levesque, // *Artificial Intelligence*. – 2014. – Vol. 212. – P. 27–35. – DOI: 10.1016/j.artint.2014.03.007
- Nilsson, N. J. Human-Level Artificial Intelligence? Be Serious! / N. J. Nilsson // *AI Magazine*. – 2005. – Vol. 26 (4). – P. 68–75.
- Pouloudi, A. Stakeholder Identification in Inter-Organizational Systems: Gaining Insights for Drug Use Management Systems / A. Pouloudi, E. A. Whitley // *European Journal of Information Systems*. – 1997. – Vol. 6, iss. 1. P. 1–14.
- Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns / M. Sallam // *In Healthcare*. – 2023. – Vol. 11, № 6. – P. 887. – MDPI. – DOI: 10.3390/healthcare11060887
- Sarker, I. H. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems / I. H. Sarker // *SN Computer Science*. – 2022. – Vol. 3, art. 158. – DOI: 10.1007/s42979-022-01043-x
- Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions / I. H. Sarker // *SN Computer Science*. – 2021. – Vol. 2 (3). – P. 1–21. – DOI: 10.1007/s42979-021-00592-x
- Shapiro, S. C. *Encyclopedia of Artificial Intelligence Second Edition* / S. C. Shapiro. – New Jersey : A Wiley Interscience Publication, 1992.
- “So What If ChatGPT Wrote It?” Multidisciplinary Perspectives on Opportunities, Challenges, and Implications of Generative Conversational AI for Research, Practice, and Policy / Y. K. Dwivedi, N. Kshetri, L. Hughes [et al.] // *International Journal of Information Management*. – 2023. – Vol. 71, art. 102642. DOI: 10.1016/j.ijinfomgt.2023.102642
- The Ten Commandments of Ethical Medical AI / H. Muller, M. T. Mayrhofer, E. B. Van Veen, A. Holzinger // *Computer*. – 2021. – Vol. 54, iss. 7. – P. 119–123. – DOI: 10.1109/MC.2021.3074263

REFERENCES

- Ivanov D.A., Ivanova M.A., Sokolov B.V. Analiz tendentsiy izmeneniya printsipov upravleniya predpriyatiyami v usloviyakh razvitiya tekhnologiy industrii 4.0 [Analysis of Trends in Changes in the Principles of Enterprise Management in the Context of the Development of Industry 4.0 Technologies]. *Tr. SPIIRAN* [Proceedings of SPIIRAS], 2018, iss. 5 (60), pp. 97-127. DOI: 10.15622/sp.60.4
- Kabalina V.I., Mondrus O.V. Kontekstualizatsiya upravleniya talantami v rossiyskikh IT-kompaniyakh [Contextualizing Talent Management in Russian IT Companies]. *Rossiyskiy zhurnal menedzhmenta* [Russian Management Journal], 2018, vol. 16, no. 1, pp. 5-36. DOI: <https://doi.org/10.21638/11701/spbu18.2018.101>
- Karpov V.E., Gotovtsev P.M., Roizenzon G.V. K voprosu ob etike i sistemakh iskusstvennogo intellekta [On the Issue of Ethics and Artificial Intelligence Systems]. *Filosofiya i obshchestvo* [Philosophy and Society], 2018, vol. 2 (87), pp. 84-105. DOI: 10.30884/jfio/2018.02.07
- Kachalov R.M., Sleptsova Yu.A. Biznes-ekosistemy v krizisnykh usloviyakh: vyyavlenie faktorov riska [Business Ecosystems Under Crisis Conditions: Identifying Risk Factors]. *Rossiyskiy zhurnal menedzhmenta* [Russian Management Journal], 2022, vol. 20, no. 2, pp. 155-171. DOI: <https://doi.org/10.21638/spbu18.2022.201>
- Kachalov R.M., Sleptsova Yu.A. Upravlenie riskom v deyatel'nosti biznes-ekosistem v period krizisov [Management in Business Ecosystems During Crises]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Ekonomika* [Journal of Volgograd State University. Economics], 2022a, vol. 24, no. 2, pp. 146-159. DOI: <https://doi.org/10.15688/ek.jvolsu.2022.2.12>
- Kachalov R.M., Sleptsova Yu.A. Eticheskie problemy primeneniya algoritmov iskusstvennogo intellekta v upravlenii predpriyatim [Ethical Problems of Using Artificial Intelligence Algorithms in Enterprise Management]. *Sistemnyy analiz v proektirovanii i upravlenii* [System Analysis in Design and Management], 2023, vol. 26(2), pp. 16-23. DOI: 10.18720/SPBPU/2/id23-77
- Kleiner G.B. Informatsionnaya teoriya faktorov upravleniya ekonomicheskimi organizatsiyami [Information Theory of Economic Organizations Management Factors]. *Rossiyskiy zhurnal menedzhmenta* [Russian Management Journal], 2022, vol. 20, no. 4, pp. 461-481. DOI: <https://doi.org/10.21638/spbu18.2022.401>
- Lisovsky A.L. Primenenie neyrosetevykh tekhnologiy dlya razrabotki sistem upravleniya [Application of Neural Network Technologies for Management Development of Systems]. *Strategicheskie resheniya i risk-menedzhment* [Strategic Decisions and Risk Management], 2020, vol. 11 (4), pp. 378-389. DOI: 10.17747/2618-947X-2020-4-378-389
- Lobacheva A.S., Sobol O.V. Etika primeneniya iskusstvennogo intellekta v upravlenii personalom [Ethics of the Application of Artificial Intelligence in Human Resource Management]. *E-Management*, 2021, vol. 4 (1), pp. 20-28. DOI: 10.26425/2658-3445-2021-4-1-20-28
- Yusupov R.M., Musaev A.A. Osobennosti otsenivaniya effektivnosti informatsionnykh sistem i tekhnologiy [Features of Assessing the Effectiveness of Information Systems and Technologies]. *Tr. SPIIRAN* [Proceedings of SPIIRAS], 2017, no. 2 (51), pp. 5-34. DOI: 10.15622/sp.51.1
- Jarrah M.H., Newlands G., Lee M.K., Wolf C.T., Kinder E., Sutherland W. Algorithmic Management in a Work Context. *Big Data & Society*, 2021, vol. 8 (2), art. 20539517211020332. DOI: 10.1177/20539517211020332
- Yang J., Chen Y.L., Por L.Y., Ku C.S. A Systematic Literature Review of Information Security in Chatbots. *Applied Sciences*, 2023, vol. 13 (11), p. 6355. DOI: 10.3390/app13116355
- Seidel S., Berente N., Lindberg A., Lyytinen K., Nickerson J.V. Autonomous Tools and Design: A Triple-Loop Approach to Human-Machine Learning. *Communications of the ACM*, 2018, vol. 62, iss. 1, pp. 50-57. DOI: 10.1145/3210753
- Chiu T.K.F. The Impact of Generative AI (GenAI) on Practices, Policies and Research Direction in Education: A Case of ChatGPT and Midjourney. *Interactive Learning Environments*, 2023. DOI: 10.1080/10494820.2023.2253861
- Clarke R. Principles and Business Processes for Responsible AI. *Computer Law & Security Review*, 2019, vol. 35, iss. 4, pp. 410-422. DOI: 10.1016/j.clsr.2019.04.007
- Collins H. *Artificial Intelligence: Against Humanity's Surrender to Computers*. Polity Press, 2018.
- Julia L. *L'intelligence artificielle n'existe pas*. Paris, Éditions First, 2019.
- Le Cun Y. *Quand la machine apprend: la révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob, 2019.
- Levesque H.J. On Our Best Behaviour. *Artificial Intelligence*, 2014, vol. 212, pp. 27-35. DOI: 10.1016/j.artint.2014.03.007
- Nilsson N.J. Human-Level Artificial Intelligence? Be Serious! *AI Magazine*, 2005, vol. 26 (4), pp. 68-75.

- Pouloudi A., Whitley E.A. Stakeholder Identification in Inter-Organizational Systems: Gaining Insights for Drug Use Management Systems. *European Journal of Information Systems*, 1997, vol. 6, iss. 1, pp. 1-14.
- Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *In Healthcare*, 2023, vol. 11, no. 6, p. 887. MDPI. DOI: 10.3390/healthcare11060887
- Sarker I.H. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, 2022, vol. 3, art. 158. DOI: 10.1007/s42979-022-01043-x
- Sarker I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2021, vol. 2 (3), pp. 1-21. DOI: 10.1007/s42979-021-00592-x
- Shapiro S.C. *Encyclopedia of Artificial Intelligence Second Edition*. New Jersey, A Wiley Interscience Publication, 1992.
- Dwivedi Y.K., Kshetri N., Hughes L., Slade E.L., Jeyaraj A., Kar A.K., Baabdullah A.M., Koohang A., Raghavan V., Ahuja M., Albanna H., Albashrawi M.A., Al-Busaidi A.S., Balakrishnan J., Barlette Y., Basu S., Bose I., Brooks L., Buhalis D., Carter L., Wright R. "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges, and Implications of Generative Conversational AI for Research, Practice, and Policy. *International Journal of Information Management*, 2023, vol. 71, art. 102642. DOI: 10.1016/j.ijinfomgt.2023.102642
- Muller H., Mayrhofer M.T., Van Veen E.B., Holzinger A. The Ten Commandments of Ethical Medical AI. *Computer*, 2021, vol. 54, iss. 7, pp. 119-123. DOI: 10.1109/MC.2021.3074263

Information About the Authors

Roman M. Kachalov, Doctor of Sciences (Economics), Professor, Chief Researcher, Central Economics and Mathematics Institute of the Russian Academy of Sciences, Prosp. Nakhimovskiy, 47, 117418 Moscow, Russian Federation, kachalov@cemi.rssi.ru, <https://orcid.org/0000-0001-5866-3390>

Yulia A. Sleptsova, Candidate of Sciences (Economics), Leading Researcher, Central Economics and Mathematics Institute of the Russian Academy of Sciences, Prosp. Nakhimovskiy, 47, 117418 Moscow, Russian Federation, julia_sleptsova@mail.ru, <https://orcid.org/0000-0001-9343-3574>

Информация об авторах

Роман Михайлович Качалов, доктор экономических наук, профессор, главный научный сотрудник, Центральный экономико-математический институт РАН, просп. Нахимовский, 47, 117418 г. Москва, Российская Федерация, kachalov1ya@yandex.ru, <https://orcid.org/0000-0001-5866-3390>

Юлия Анатольевна Слепцова, кандидат экономических наук, ведущий научный сотрудник, Центральный экономико-математический институт РАН, просп. Нахимовский, 47, 117418 г. Москва, Российская Федерация, julia_sleptsova@mail.ru, <https://orcid.org/0000-0001-9343-3574>