УДК 65.01
DOI: 10.24412/2312-6647-2026-147-163-187

# РИСКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ МЕНЕДЖМЕНТА: КЛАССИФИКАЦИЯ, УПРАВЛЕНИЕ, KPI

**Ольга Сергеевна Елкина**

Российская академия народного хозяйства и государственной службы при Президенте РФ, Северо-Западный институт управления, Санкт-Петербург, Россия,
phdelkina@mail.ru, https://orcid.org/0000-0003-4952-1512

***Аннотация.*** Стремительное внедрение технологий искусственного интеллекта (ИИ) в бизнес-процессы качественно трансформирует природу управленческих рисков, порождая новые стратегические, операционные, этические и репутационные угрозы. Автономность и непрозрачность ИИ-систем требуют разработки специализированных подходов к их управлению, интегрированных в корпоративную стратегию. Цель исследования — систематизация рисков ИИ для менеджмента и разработка системы управления этими рисками, включая подбор оптимальных ключевых показателей эффективности (KPI). В основе работы лежит вторичный анализ данных отчетов, опросов и исследовательских публикаций. Использованы методы казуистического, количественного и системного анализа, а также нормативный подход для формирования управленческих рекомендаций. Исследование выявило и структурировало семь ключевых категорий рисков ИИ: стратегические, операционные, этико-правовые, организационные, финансово-инвестиционные, риски для управленческих ролей и репутационные. Для каждой категории предложены конкретные управленческие практики и система KPI, связывающая технические риски с бизнес-последствиями. Научная ценность работы заключается в комплексной систематизации ИИ-рисков для менеджмента. Практическая значимость состоит в предоставлении руководителям структурированного инструментария для интеграции управления рисками ИИ в корпоративную стратегию и операционную деятельность, что является критическим фактором устойчивого развития в условиях цифровой трансформации.

***Ключевые слова:*** искусственный интеллект, риски искусственного интеллекта, управление рисками, KPI для менеджмента.

UDC 65.01
DOI: 10.24412/2312-6647-2026-147-163-187

# RISKS OF ARTIFICIAL INTELLIGENCE FOR MANAGEMENT: CLASSIFICATION, MANAGEMENT, KPI

**Olga Sergeevna Elkina**

The Russian Presidential Academy of National Economy
and Public Administration, North-West Institute of Management,
Saint Petersburg, Russia,
phdelkina@mail.ru, https://orcid.org/0000-0003-4952-1512

***Abstract.*** The rapid introduction of artificial intelligence (AI) technologies into business processes qualitatively transforms the nature of management risks, generating new strategic, operational, ethical and reputational threats. The autonomy and transparency of AI systems require the development of specialized approaches to their management, integrated into the corporate strategy. The purpose of the study — systematization of AI risks for management and development of a risk management system, including the selection of optimal key performance indicators (KPIs). The work is based on the secondary analysis of data from reports, surveys and research publications. The methods of statistical, quantitative and systematic analysis, as well as a normative approach for the formation of management recommendations are used. The study identified and structured seven key categories of AI risks: strategic, operational, ethical, organizational, financial and investment, risks for managerial roles, and reputational. Specific management practices and a KPI system are proposed for each category, linking technical risks with business consequences. The scientific value of the work lies in the comprehensive systematization of AI risks for management. The practical value is to provide managers with structured tools for integrating AI risk management into corporate strategy and operational activities, which is a critical factor for sustainable development in the context of digital transformation.

***Keywords:*** artificial intelligence, artificial intelligence risks, risk management, KIP for management.

## Введение

Необходимость систематического изучения рисков искусственного интеллекта в прикладном менеджменте обосновывается не только быстротой внедрения решений ИИ в ключевые бизнес-процессы, но и качественной трансформацией характера управленческих рисков: автономность и непрозрачность современных моделей порождают новые типы операционных, репутационных и этических угроз, требующих специализированных методов их идентификации, оценки и смягчения. Так, Стюарт Рассел в монографии «Совместимость с человеком» [1] подчеркивает, что рост автономности ИИ делает централизованные предпосылки контроля недостаточными и требует переосмысления задач безопасности и сопоставления целей системы с интересами людей.

Эмпирические исследования дополняют эту картину: анализ инцидентов и внутренней документации платформ показывает, что ИИ-механизмы, ориентированные на максимизацию вовлеченности или другие коммерческие метрики, способны непреднамеренно усиливать вредные эффекты в виде распространения дезинформации, ухудшения психического здоровья целевых групп и т. д. Это было проиллюстрировано и систематизировано в обзорах и материалах MIT Technology Review [2]. Данные примеры свидетельствуют о том, что менеджеры сталкиваются с рисками, которые не сводятся к багам, и требуют междисциплинарного управления.

С позиций нормативно-правовой и институциональной ответственности ключевые работы по подотчетным алгоритмам указывают на необходимость расширения рамок аудита и корпоративной отчетности: Дж. Кролл и соавторы в статье «Подотчетные алгоритмы» [3] предлагают юридические и организационные механизмы, призванные связать техническую валидацию моделей с управленческими процессами принятия решений и ответственностью. Это подчеркивает, что риски ИИ не только технически, но и институционально уже сформированы, и, следовательно, решение по управлению ими лежит в плоскости координации между техническими специалистами, юристами и менеджерами.

Проведенные критические исследования, в частности монография Шошаны Зубофф [4], демонстрируют социальные и распределительные аспекты алгоритмических рисков: концентрация данных и непрозрачные практики моделирования могут усиливать асимметрию власти и создавать системные несправедливости, что трансформирует традиционные представления о корпоративной ответственности и требует включения этических оценок в стратегическое управление.

Развитость инструментов управления рисками ИИ на сегодняшний день остается смешанной: появились зрелые фреймворки для оценки и управления рисками (включая NIST AI Risk Management Framework), однако их трансляция в повседневные управленческие практики осложнена отсутствием единых процедур аудита, недостаточной интеграцией рисков ИИ в корпоративные процедуры управления рисками и дефицитом подготовленных управленцев. В результате исследовательская повестка должна быть направлена на создание воспроизводимых процедур аудита, метрик подотчетности и программ повышения организационной готовности, которые объединяли бы технические и управленческие компетенции.

Таким образом, научная и практическая аргументация указывает на критическую важность дальнейших междисциплинарных исследований: для менеджмента это означает переход от эпизодического реагирования к системной интеграции управления ИИ-рисками в стратегические и операционные практики организации.

В этой связи основной целью исследования является систематизация ИИ-рисков для менеджмента и разработка системы управления ими.

Задачи исследования заключаются в систематизации ИИ-рисков для менеджмента, разработке системы управления этими рисками и подбор оптимальных KPI, стимулирующих менеджмент к управлению этими ИИ-рисками.

## Методы исследования

Мы использовали метод вторичного анализа отчетов внешних организаций, результатов опросов и исследовательских публикаций для описания и систематизации существующих ИИ-рисков. Этот метод позволил систематизировать уже накопленные эмпирические данные и представить их в контексте менеджерского риска. В работе применен казуистический подход, позволивший проиллюстрировать, как абстрактные риски приобретают практическую форму. Количественный метод анализа статистики позволил проанализировать количественные эмпирические данные. Метод системного анализа и концептуализации способствовал структурированию рисков по категориям. При описании рекомендаций для менеджмента применен нормативный подход: выстроены желаемые практики и процедуры, основанные на этических, правовых и организационных принципах. Такой набор методов позволил обеспечить всесторонний обзор рисков ИИ для менеджмента и предложить обоснованные управленческие решения.

## Результаты исследования

В эпоху стремительного распространения технологий искусственного интеллекта управление организациями становится все более зависимым от ИИ-систем, а это порождает комплекс новых рисков для менеджмента. Несмотря на очевидные выгоды, связанные с автоматизацией процессов, возможностью обработки аналитики, повышением эффективности принятия решений, актуальность вопросов управления рисками ИИ возрастает. Менеджменту необходимо не только реализовывать возможности ИИ, но и учитывать потенциал нежелательных последствий с точки зрения стратегии, операций, этики, организации и репутации.

### Стратегические риски

Во-первых, использование ИИ может привести к утрате стратегического контроля со стороны менеджмента. Архитектура нейронной сети, внедренная для поддержки решений (например, прогнозирование спроса, распределение ресурсов или динамическое ценообразование), в ряде случаев становится черным ящиком, что снижает способность управленцев видеть общую картину и выстраивать долгосрочную стратегию.

На практике такая ситуация приводит к тому, что менеджер по продажам не понимает, почему ИИ предсказывает падение спроса на ключевой продукт. Финансовый директор не может объяснить совету директоров логику, по которой ИИ рекомендовал свернуть перспективный, но пока убыточный проект. Компания начинает плыть по течению, которое определяется некой обученной моделью, и теряет способность к стратегическому маневру в своей деятельности.

Стратегия — это не просто реакция на данные, это видение, интуиция и понимание контекста, которыми машина не обладает. Слепое доверие к черному ящику ведет к стратегическому параличу, когда компания не может действовать без одобрения ИИ-систем. Упускаются возможности, потому что нейронная сеть не всегда может учитывать слабые сигналы или нарождающиеся тренды, которые видит опытный менеджер. При изменении рыночных условий (как во время пандемии) модель, обученная на старых данных, может давать губительные рекомендации, и если менеджеры не смогут вовремя это заметить, то следование подобным рекомендациям может привести к катастрофическим ошибкам.

Во-вторых, существует риск несоответствия целей ИИ корпоративным ценностям и стратегическим ориентирам. Например, обученная модель может оптимизировать краткосрочную прибыль в ущерб лояльности клиентов, репутации или инновационной способности организации.

Типичная ситуация — ИИ-системы кол-центров, которые оценивают операторов по скорости звонка. Они оптимизируют метрику «время разговора», но убивают клиентский сервис, поскольку операторы начинают спешить и сбрасывать сложных клиентов. Краткосрочная эффективность достигается ценой долгосрочной лояльности. Это прямая угроза бизнес-модели и бренду. В результате компания может незаметно для себя начать технологически обосновывать неэтичные практики или разрушать свои конкурентные преимущества. Например, инновации требуют терпимости к неудачам и долгосрочных инвестиций, что противоречит процессу оптимизации квартальных финансовых показателей.

Статистика подтверждает, что компании все чаще интерпретируют ИИ не только как возможность, но и как существенный риск. По данным исследования, проведенного Financial Times, среди крупнейших корпораций США доля тех, которые упомянули ИИ как фактор риска, выросла с 9 % в 2022 г. до 56 % в 2025-м[1]. Это свидетельствует о понимании топ-менеджментом того, что ИИ может стать стратегической угрозой, если ему не уделяется должного внимания.

[1] Biggest US companies warn of growing AI risk // Financial Times. 2024. August 17. URL: https://www.ft.com/content/5ee96d38-f55b-4e8a-b5c1-e58ce3d4111f (дата обращения: 30.11.2025).

### Операционные риски

Ключевые операционные риски связаны с технологическими сбоями, ошибками ИИ-систем, а также с несоответствием ИИ существующим бизнес-процессам, инфраструктуре и рыночной ситуации.

В отчете крупнейшей международной сети аудиторских и консалтинговых фирм, возглавляемой компанией KPMG[2], представлены результаты исследования по трем главным категориям рисков, которые компания активно пытается контролировать: целостность данных, статистическая обоснованность и точность модели.

Целостность данных определяет основу для корректных выборок и признаков, под которой KPMG понимает степень достоверности, непротиворечивости и полноты данных, используемых для обучения, тестирования и эксплуатации моделей. Нарушение целостности — это не только технический дефект (например, ошибки ввода или поврежденные записи), но и управленческий риск: данные могут быть смещенными, неполными, устаревшими или собранными без надлежащей валидации источников. В отчете KPMG 2023 г.[3] отмечается, что 67 % организаций сталкивались с проблемами качества и консистентности данных при внедрении ИИ, а 54 % признали, что отсутствие формальных процедур происхождения данных препятствует надежному аудиту моделей. С точки зрения менеджмента это означает, что система ИИ не может быть надежнее своих данных. Целостность данных — основа для устойчивости результатов при изменении источников, воспроизводимости аналитики и доверия к отчетности и объяснимости модели.

Статистическая обоснованность указывает на корректность выбора методов, достаточность объема и репрезентативность данных, а также на то, насколько статистические выводы модели являются значимыми и устойчивыми (можно ли вообще доверять метрикам точности). В своем отчете KPMG подчеркивает, что многие ИИ-системы страдают от псевдостатистической устойчивости — внешне корректных метрик при нарушении предпосылок статистической модели. Например, несоблюдение независимости наблюдений, чрезмерный дисбаланс классов или неправильная валидация приводят к ложному чувству надежности. По данным KPMG, около 48 % компаний не проводят регулярную перекалибровку моделей и не проверяют статистическую устойчивость прогнозов на новых данных, что повышает риск некорректных решений при изменении внешней среды (например, макроэкономических условий). С управленческой точки зрения статистическая обоснованность — это гарантия того, что ИИ действительно извлекает закономерности, а не подгоняет результаты под случайные шумы.

---

[2] Responsible AI and the challenge of AI risk // KMPG. 2023. July 11. URL: https://kpmg.com/be/en/home/insights/2023/07/lh-responsible-ai-and-the-challenge-of-ai-risk.html (дата обращения: 24.12.2025).

[3] Там же.

Точность модели отражает ее способность адекватно предсказывать или классифицировать события в реальных условиях эксплуатации. В отчетах KPMG точность рассматривается не только как статистическая метрика, но и как управленческая характеристика — степень соответствия результата ожиданиям бизнеса и регуляторным требованиям. Однако KPMG указывает, что компании часто переоценивают важность «сырых метрик точности» в ущерб системной устойчивости. В их исследовании 41 % руководителей заявили, что основным критерием оценки ИИ остается точность, при том что менее 30 % учитывают метрики объяснимости и справедливости, что ведет к асимметричному контролю рисков.

Таким образом, точность без контроля целостности данных и статистической корректности становится ложно высокой: модель может демонстрировать хорошие показатели на тестовом наборе, но быть нерелевантной в реальной среде.

Согласно исследованию Дж. Поста[4], за 2025 г. доля руководителей, рассматривающих ИИ как риск, выросла с 5 до 11 %. Основными рисками, которые беспокоят менеджеров, являются ошибки и галлюцинации с использованием искусственного интеллекта (34 %), повышенная угроза нарушения конфиденциальности или утечек данных (33 %), а также ответственность или юридическое обоснование неправомерного использования искусственного интеллекта (31 %). Согласно результатам опроса, многие компании обновляют свои системы управления рисками и пересматривают свои стратегии, чтобы учесть соображения искусственного интеллекта, пересматривая протоколы кибербезопасности, усиливая политику конфиденциальности данных и обеспечивая соответствие нормативным требованиям.

Такие данные указывают на то, что ИИ-инструменты могут породить ошибки в прогнозировании, автоматизации или контроле, что, в свою очередь, может привести к неверным управленческим решениям.

### Этические и правовые риски

Включают системную предвзятость, воспроизводимую ИИ, нарушение приватности и защиты данных, а также неопределенность в вопросах ответственности.

Согласно обзору Harvard Law School Corporate [5], примерно 20 % компаний сообщили о том, что их ИИ-модели, либо ИИ-системы, либо обучающие методики могли быть дефектными или вызывать социальный вред, включая дискриминационные или некорректные выводы, нарушения приватности и непредсказуемое воздействие на заинтересованные стороны. При этом, как отмечает

[4] *Post J.* Leaders increasingly concerned about AI adoption risks // Risk management magazine. 2025. July 1. URL: https://www.rmmagazine.com/articles/article/2025/07/01/leaders-increasingly-concerned-about-ai-adoption-risks (дата обращения: 09.11.2025).

тот же источник, лишь небольшая доля компаний имеет формализованные программы управления рисками ИИ, а среди существующих программ около половины не способны эффективно устранить или смягчить риски, связанные с этическими и юридическими аспектами использования ИИ.

Системная предвзятость, воспроизводимая ИИ, возникает, когда обучающие данные или методы построения модели приводят к систематическим искажениям, которые дискриминируют определенные группы пользователей по признакам пола, возраста, этничности, социального статуса или иных характеристик. По сути, предвзятость ИИ-модели — это не просто техническая ошибка, а социально-правовой риск, способный нарушить принципы справедливости и равенства, которые закреплены в нормативных актах.

Исследования Harvard Law School Corporate показывают, что до 38 % крупных компаний уже сталкивались с обвинениями в предвзятости ИИ-моделей, особенно в сферах найма персонала, кредитного скоринга и таргетированной рекламы. Примером служит кейс Amazon, где рекрутинговая система занижала рейтинг кандидаток-женщин из-за исторического перекоса в обучающих данных. Этический аспект состоит в том, что даже непреднамеренные искажения могут воспроизводить системную дискриминацию, а юридический — в том, что подобные случаи подпадают под нормы антидискриминационного законодательства и создают основу для судебных исков.

ИИ-системы требуют значительных объемов данных, включая персональные и чувствительные. Это создает риски нарушения приватности, неправомерного использования данных и несоблюдения требований законодательства, в частности Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных»[5].

Harvard Law Scool Forum of Corporate Governance [5] указывает, что более 60 % компаний испытывают сложности с обеспечением прозрачности цепочек обработки данных и часто не имеют четких процедур согласия пользователей на использование их данных для обучения моделей.

Этические последствия касаются информированного согласия и цифрового суверенитета личности, тогда как правовые — возможности наложения штрафов, санкций или судебных исков. В 2021–2023 гг. крупные технологические компании (в том числе и Google) выплачивали штрафы в сотни миллионов долларов за несанкционированную обработку данных для обучения моделей рекламы и распознавания изображений.

В российской практике аналогичные риски проявляются через необходимость соблюдения требований Роскомнадзора к локализации и обработке персональных данных, а также через недавние инициативы Минцифры РФ, направленные на регулирование обучения ИИ-моделей на отечественных датасетах.

Ключевой юридический вызов в области ИИ — это распределение ответственности за действия или решения, принятые моделью. В традиционных управленческих схемах ответственность закрепляется за конкретным лицом

[5] URL: https://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 17.11.2025).

или подразделением, но в системах ИИ решение формируется распределенно — в результате работы алгоритма, обученного на множестве источников данных.

Отчет Harvard Law School Corporate фиксирует, что около 45 % организаций не имеют четко определенного механизма фиксации ответственности в случае ошибок ИИ и только 12 % внедрили внутренние процедуры регистрации инцидентов ИИ. Это создает разрыв между фактическим воздействием модели и юридическим субъектом, способным за него отвечать.

В международной практике такие пробелы уже вызывают необходимость нормативных реформ. В частности, проект EU AI Act (2024)[6] прямо требует наличия ответственного лица (AI Officer) и механизмов документирования всех решений модели. В корпоративном управлении это трансформируется в создание специализированных комитетов по управлению рисками ИИ.

С точки зрения менеджмента этические и правовые риски ИИ подрывают корпоративное доверие, репутационный капитал и устойчивость к регуляторным изменениям.

Проблема заключается в том, что большинство компаний по-прежнему рассматривают эти риски как внешние — относящиеся к юридическим департаментам, а не к стратегическому управлению. По данным Harvard Law School Corporate, отсутствие интегрированного подхода приводит к тому, что только 25 % советов директоров регулярно обсуждают этические вопросы ИИ на уровне корпоративной повестки.

Это означает, что менеджменту предстоит учитывать не только корректность технических решений, но и их влияние на социальные, правовые и этические аспекты: дискриминация, частные данные, вопросы ответственности.

## Организационные риски

Организационные риски, связанные с внедрением искусственного интеллекта, представляют собой системную угрозу социально-технической архитектуре предприятия. В отличие от операционных сбоев, эти риски подрывают саму человеческую и культурную основу, необходимую для успешной интеграции технологий. Их можно концептуализировать как три взаимосвязанных вызова: человеко-машинный разрыв в доверии, утрата организационного знания и кризис управленческой компетентности.

Первым и наиболее фундаментальным вызовом является человеко-машинный разрыв в доверии, который проявляется в форме сопротивления персонала. Это происходит, когда автоматизация на основе ИИ воспринимается сотрудниками не как инструмент улучшения их деятельности, а как прямая угроза их профессиональной автономии или занятости. Это провоцирует глубокую психологическую реакцию: происходит снижение мотивации и лояльности сотрудников.

[6] EU AI Act. URL: https://artificialintelligenceact.eu/the-act/ (дата обращения: 25.12.2025).

Но более сложной проблемой становится пассивное неприятие, при котором сотрудники формально следуют предписаниям системы, но лишаются стимула к проявлению инициативы и критического мышления. Это создает скрытую стоимость внедрения ИИ: организация теряет способность к неформальной адаптации и оперативной коррекции процессов, которая традиционно обеспечивалась человеческой интуицией и спонтанным сотрудничеством.

Разрыв в доверии усугубляет вторую ключевую проблему — системную потерю неявного знания. Неявное знание, в отличие от явных знаний (представленных инструкциями и регламентами), представляет собой личный, контекстуально обусловленный и часто невербализуемый опыт (интуиция, опыт и т. д.). Именно оно позволяет опытному менеджеру почувствовать назревающий конфликт в команде или специалисту принять нестандартное решение в уникальной ситуации. Активное внедрение ИИ-систем, оптимизированных под формализуемые метрики, неявно дискредитирует ценность такого знания. Происходит «процессуальное вытеснение»: сотрудники, следуя предписаниям ИИ-моделей, перестают накапливать и применять собственный эмпирический опыт. В долгосрочной перспективе это ведет к обеднению организационного интеллекта и снижению гибкости компании, которая оказывается неспособной адекватно реагировать на кризисы, не прописанные в сценариях работы ИИ.

Наконец, третьим системным риском является кризис управленческой компетентности, который усугубляет две предыдущие проблемы. Статистические данные, согласно которым значительный процент менеджеров не имеет опыта взаимодействия с базовыми инструментами ИИ и не проходил соответствующего обучения, указывают на возникновение критического разрыва между ответственностью и пониманием. Менеджер, лишенный цифровой грамотности в области ИИ, не способен ни корректно поставить задачу для ИИ-системы, ни оценить качество ее работы, ни, что важнее, распознать ситуацию, в которой применение ИИ неуместно. Это создает порочный круг: некомпетентное руководство внедряет системы, порождающие сопротивление сотрудников и вытесняющие неявное знание, а затем оказывается не в состоянии управлять последствиями этих процессов из-за того же дефицита понимания. Таким образом, дефицит навыков на уровне управления не просто является отдельным риском, но выступает катализатором, усиливающим всю систему организационных угроз, что в конечном итоге ведет к глубокой институциональной нестабильности.

Согласно опросу, проведенному Times[7], более трети британских менеджеров никогда не пользовались, например, инструментом ИИ ChatGPT и 86 % не получали формального обучения ИИ. По данным аналитического центра макроэкономических исследований Сбера, российские компании находятся на начальной стадии внедрения ИИ. Фокус смещен в сторону пилотных

[7] *Tyler R.* A third of managers in the UK have never used AI // Times. 2024. August 21. URL: https://www.thetimes.com/business-money/entrepreneurs/article/a-third-of-managers-in-uk-have-never-used-ai-enterprise-network-gm69j5flh (дата обращения: 09.12.2025).

проектов и экспериментов, а не на масштабирование. Это означает, что даже в компаниях, где есть ИИ-инициативы, большинство линейных и средних менеджеров могут не иметь к ним системного доступа. Использование остается уделом узких специалистов или топ-менеджмента.

Это подчеркивает пробелы компетенций, что само по себе является риском: управление ИИ без соответствующей подготовки усилит потенциальную нестабильность организации.

### Финансово-инвестиционные риски

Анализ экономических аспектов внедрения искусственного интеллекта выявляет фундаментальный парадокс, при котором прогнозируемость затрат контрастирует с высокой неопределенностью возврата инвестиций (ROI). Данное противоречие коренится в самой природе ИИ-проектов, где первоначальные расходы на технологии и инфраструктуру являются лишь видимой частью айсберга. Существенные, но часто скрытые издержки возникают на последующих этапах: интеграция с имеющимися системами требует значительных ресурсов, а непрерывный мониторинг, обновление моделей и обеспечение их безопасности формируют постоянную операционную нагрузку. При этом многие организации принимают инвестиционные решения в условиях существенного информационного разрыва, не имея четких метрик для оценки будущей эффективности и без системного понимания сопутствующих рисков.

В этой связи показателен прогноз динамичного роста глобального рынка ИИ в сфере управления рисками. Цифры, свидетельствующие о потенциальном увеличении объема рынка до 15,5 млрд долл. к 2028 г.[8], отражают растущее коллективное признание ценности технологии. Однако подобные макроэкономические тенденции не отменяют индивидуальных трудностей на уровне отдельной компании. Напротив, они создают прессинг, подталкивая организации к догоняющему внедрению без выстраивания адекватной системы управления.

Именно это противоречие между внешними ожиданиями и внутренней неподготовленностью порождает феномен «пилотной ловушки». В такой ситуации проект успешно проходит экспериментальную стадию, демонстрируя определенный потенциал, но оказывается неспособным к масштабированию и интеграции в основные бизнес-процессы. В результате организация сталкивается с ситуацией, когда постоянные инвестиции не приводят к сопоставимому росту ценности, а первоначальные вложения превращаются в безвозвратные затраты. Таким образом, высокая неопределенность ROI является не столько внешним рыночным условием, сколько следствием недостаточной системности в управлении самим процессом внедрения и эксплуатации технологий искусственного интеллекта.

---

[8] *Linder J.* AI in the risk management industry statistics // GITNUX. REPORT 2025. URL: https://gitnux.org/ai-in-the-risk-management-industry-statistics (дата обращения: 07.10.2025).

### Риски для управленческих ролей

Интеграция искусственного интеллекта в управленческие процессы инициирует структурную трансформацию традиционной роли менеджера. Данная трансформация характеризуется не просто автоматизацией рутинных операций, но фундаментальным перераспределением когнитивных функций между человеком и ИИ-системой. Функции, традиционно составлявшие основу управленческого труда, такие как анализ больших массивов данных, оперативное планирование и мониторинг исполнения задач, постепенно делегируются интеллектуальным системам. Это смещает фокус деятельности менеджера с обработки информации и контроля в сторону интерпретации смыслов, стратегического выбора и управления контекстом. Однако подобное перераспределение несет в себе риск дегуманизации управленческой практики, когда решения, основанные исключительно на интерпретации решений, принимаемых нейронной сетью, вытесняют ценностно-окрашенное суждение, формирующееся в рамках человеческого опыта.

Смещение фокуса в сторону ИИ-эффективности закономерно приводит к маргинализации социально-психологических аспектов управления. В ситуациях совместного принятия решений человеком и ИИ происходит снижение значимости таких качеств, как социальная чувствительность и эмпатия. Автономные ИИ-системы, будучи лишенными способности к интуитивному пониманию нюансов межличностных отношений, предлагают решения, оптимизированные под формализуемые метрики, но не учитывающие морально-психологический климат в коллективе. Когда управленец начинает в значительной степени полагаться на такие рекомендации, это может привести к снижению доверия и ощущению обезличенности взаимодействия у сотрудников.

Кумулятивный эффект этого процесса оказывает глубокое влияние на организационную культуру в целом. Постепенно вытесняя из управленческой практики неформальные, основанные на эмпатии и личном опыте взаимодействия, организация рискует трансформировать свою внутреннюю социальную систему. Коммуникация приобретает все более инструментальный характер, а способность к неформальной адаптации и коллективной поддержке, столь важная в ситуациях неопределенности, снижается. Таким образом, технологическая трансформация роли менеджера, усиливая его аналитические возможности, одновременно ставит перед организацией новую комплексную задачу — сохранения человекоцентричности в условиях управления с ИИ-системами, что требует осознанного проектирования гибридных человеко-машинных систем принятия решений.

### Репутационные риски

Репутационные риски, возникающие в контексте применения искусственного интеллекта, представляют собой системную угрозу нематериальным

активам организации, формируемым доверием заинтересованных лиц (поставщиками, клиентами и т. д.). В отличие от операционных сбоев, репутационный ущерб обладает кумулятивным характером и способностью к лавинообразной эскалации в публичном поле.

Ключевым источником такой угрозы является этико-правовая неопределенность, присущая ряду решений ИИ. Системная предвзятость, воспроизводимая ИИ, демонстрирующая дискриминационные паттерны в результате смещенных данных, системы, нарушающие приватность пользователей в процессе обучения, или автономные решения, приводящие к существенным ошибкам, немедленно попадают в фокус общественного внимания и регуляторного давления. Подобные инциденты интерпретируются обществом не как технические несовершенства, а как следствие системных сбоев в корпоративной этике и управленческой культуре. Это наносит урон фундаментальному отношению между компанией и ее клиентами, инвесторами и сотрудниками, основанному на ожидании справедливости, прозрачности и ответственности.

Растущее осознание данной уязвимости находит свое отражение в корпоративной отчетности. Тот факт, что все большее число компаний идентифицируют ИИ в качестве существенного фактора риска в своих годовых отчетах, сигнализирует о сдвиге в восприятии технологий руководством. Из инструмента оптимизации ИИ трансформируется в потенциальный источник стратегической нестабильности. Эта практика отражает не только стремление соблюсти регуляторные требования, но и попытку управлять ожиданиями рынка, предвосхищая потенциальные кризисы. Таким образом, проактивное раскрытие информации о рисках, связанных с ИИ, становится элементом стратегического управления репутацией, направленным на смягчение последствий возможных инцидентов и демонстрацию зрелого подхода к технологическим инновациям. В конечном счете в цифровую эпоху репутационная устойчивость организации становится не менее важным активом, чем ее технологическая оснащенность, и требует не менее системного управления.

## Практики, минимизирующие риски

Для минимизации перечисленных рисков рекомендуется внедрять следующие практики:

1. ***Разработка интегрированной стратегии управления ИИ-рисками***, включающая идентификацию, анализ, оценку и управление рисками на всех уровнях.

Такой подход предполагает создание целостной системы, обеспечивающей сквозное управление рисками на всех организационных уровнях — от операционного до стратегического.

Фундаментом данной системы выступает процесс идентификации рисков, который должен охватывать не только технические аспекты функционирования

ИИ-систем, но и их потенциальное воздействие на бизнес-процессы, человеческие ресурсы и внешнюю среду. Последующий качественный и количественный анализ выявленных рисков позволяет определить их вероятностные характеристики и масштаб потенциального воздействия, формируя объективную основу для расстановки приоритетов.

На основе результатов анализа осуществляется комплексная оценка рисков, интегрирующая технические метрики с бизнес-критериями приемлемости. Этот этап позволяет дифференцировать риски по степени их допустимости и определить те из них, которые требуют незамедлительных управленческих вмешательств.

Ключевым элементом стратегии становится разработка и реализация целевых мероприятий по управлению выявленными рисками, которые могут включать меры по их снижению, передаче, принятию или исключению. Эффективность данных мероприятий обеспечивается за счет создания распределенной системы ответственности, где управленческие функции делегируются соответствующим структурным подразделениям в соответствии с их компетенцией.

Неотъемлемым свойством интегрированной стратегии является ее итеративный характер, предполагающий постоянный мониторинг риск-ландшафта и адаптацию управленческих практик к изменяющимся внешним и внутренним условиям. Это обеспечивает устойчивость системы управления рисками в условиях динамичного развития ИИ-технологий и соответствующей трансформации регуляторных требований.

2. ***Утверждение эффективной системы управления данными:*** обеспечение качества, целостности и обоснованности данных, на которых строятся модели. Такая система представляет собой критический императив для обеспечения надежности и валидности искусственного интеллекта. Она должна обеспечивать соблюдение принципов качества, целостности и обоснованности данных на всех этапах жизненного цикла модели — от сбора и обработки до эксплуатации и мониторинга.

Качество данных определяется их точностью, полнотой и согласованностью, что напрямую влияет на репрезентативность и предсказательную способность ИИ-моделей. Недостатки в этом аспекте, такие как шумы, пропущенные значения или систематические смещения, способны не только снизить эффективность данной модели, но и воспроизвести или усилить существующие в данных предубеждения, приводя к дискриминационным или социально вредным результатам.

Целостность данных подразумевает их защищенность от несанкционированных изменений, а также устойчивость к преднамеренным искажениям. Нарушение целостности ставит под угрозу достоверность выводов модели и может быть использовано для скрытого манипулирования ее работой, что создает операционные и репутационные риски для организации.

Наконец, обоснованность данных относится к их релевантности и адекватности поставленной задаче. Даже технически безупречные данные, собранные без учета контекста или целевого назначения модели, могут привести к построению корректных с математической точки зрения, но бессмысленных или ошибочных с практической точки зрения интеллектуальных систем. Это требует тесной интеграции предметных экспертов в процессы формирования и валидации данных.

Создание комплексной системы управления данными является не технической задачей, а стратегической необходимостью, обеспечивающей основу для ответственного и эффективного использования искусственного интеллекта.

3. ***Прозрачность и объяснимость ИИ-моделей.*** Менеджмент должен понимать, как принимаются решения ИИ-системами, и быть способен объяснять их заинтересованным сторонам.

Данное требование обусловлено не только операционной необходимостью, но и стратегической задачей поддержания доверия со стороны всех заинтересованных лиц. Прозрачность подразумевает доступность информации о принципах функционирования системы, используемых данных и ограничениях модели, тогда как объяснимость относится к способности содержательно интерпретировать конкретные решения, выдаваемые ИИ.

Для управленческого персонала понимание логики принятия решений искусственным интеллектом является необходимым условием для осуществления содержательного контроля над автоматизированными процессами. Когда менеджер способен реконструировать причинно-следственные связи, приведшие к тому или иному результату, он сохраняет возможность критической оценки и корректировки действий системы. Это особенно значимо в ситуациях, когда ИИ рекомендации вступают в противоречие с экспертной оценкой или этическими нормами организации.

Способность менеджмента объяснять принципы работы ИИ-моделей ключевым заинтересованным сторонам, включая регуляторов, клиентов и сотрудников, формирует основу для публичной легитимации использования ИИ. В условиях растущего общественного внимания к вопросам цифровой этики, организации, демонстрирующие внимание к объяснимости, получают существенные конкурентные преимущества в области репутационного капитала.

Техническая реализация объяснимости требует внедрения специализированных методик интерпретации моделей, которые позволяют визуализировать весомость различных факторов в итоговом решении. Однако следует признать, что существует объективный компромисс между точностью сложных моделей и степенью их интерпретируемости, что порождает методологическую дилемму при проектировании систем искусственного интеллекта.

Инвестиции в прозрачность ИИ-систем следует рассматривать как необходимые меры по снижению репутационных, правовых и операционных рисков, связанных с внедрением ИИ-решений.

4. ***Обучение и развитие компетенций.*** Менеджеры и сотрудники должны получать необходимое обучение по ИИ, его возможностям и рискам (учитывая факты, что многие не имеют формальной подготовки).

Необходимость реализации программ обучения и развития компетенций в области ИИ обусловлена преодолением существенного разрыва между динамичным развитием технологий и существующим уровнем подготовки управленческих кадров и сотрудников.

Содержательное наполнение образовательных программ должно выходить за рамки операционного освоения инструментария, охватывая формирование системного понимания возможностей и ограничений искусственного интеллекта. Особое значение приобретает развитие критического мышления, позволяющего осуществлять содержательную интерпретацию результатов работы ИИ-моделей и выявлять потенциальные риски их применения. Менеджеры должны приобрести компетенции, необходимые для постановки задач системам ИИ, оценки адекватности их выводов и интеграции алгоритмических решений в существующие бизнес-процессы.

При этом образовательные инициативы должны учитывать дифференциацию профессиональных ролей и уровней взаимодействия с технологией. Если для технических специалистов приоритетом является глубокое понимание архитектурных особенностей и методов разработки ИИ-моделей, то для управленческого персонала ключевое значение приобретает формирование ИИ-интуиции — способности прогнозировать поведение системы в условиях неопределенности и оценивать ее воздействие на организационные структуры.

Интеграция знаний об искусственном интеллекте в систему корпоративного обучения способствует преодолению психологических барьеров и сопротивления персонала, возникающих при внедрении новых технологий. Когда сотрудники понимают принципы функционирования ИИ и границы его компетенции, формируется основа для конструктивного сотрудничества человека и машины, при котором технология воспринимается как инструмент усиления человеческих способностей, а не их замены.

Инвестиции в развитие компетенций в области ИИ непосредственно влияют на эффективность цифровой трансформации и долгосрочную конкурентоспособность компании в условиях становления экономики, основанной на знаниях.

5. ***Интеграция человека и машины.*** Применение ИИ должно дополнять, а не замещать управленческое суждение, сохранять роль «человека-в-контроле».

Реализация потенциала ИИ в управленческой практике требует преодоления парадигмы автоматизации в пользу модели гибридного интеллекта, где технологические системы функционируют в качестве когнитивных усилителей принятия решений человеком. Ключевым принципом такой интеграции становится сохранение за ним роли конечного арбитра, осуществляющего содержательный контроль над ИИ-рекомендациями и несущего ответственность за итоговые управленческие решения.

Эффективное взаимодействие в системе «человек – машина» предполагает не механическое противопоставление интуитивного опыта и ИИ-выводов, а их содержательный синтез. Управленческое суждение, обогащенное результатами работы искусственного интеллекта, приобретает способность оперировать не только качественными оценками, но и количественно верифицированными паттернами, выявленными в многомерных данных. При этом критически важным остается сохранение способности человека учитывать контекстуальные факторы, этические императивы и стратегические приоритеты, которые остаются за пределами формализуемых параметров.

Сохранение роли «человека-в-контроле» обеспечивает необходимый баланс между эффективностью технологических решений и адаптивностью человеческого мышления. Практическая реализация данного подхода требует проектирования интерфейсов взаимодействия, обеспечивающих прозрачность работы ИИ-систем и содержательную интерпретацию их выводов. Это позволяет менеджеру не просто принимать или отвергать рекомендации системы, но и понимать их логические основания, выявлять потенциальные смещения и корректировать параметры анализа в соответствии с динамикой внешней среды.

Таким образом, интеграция ИИ в управленческие процессы должна рассматриваться не как техническая задача делегирования полномочий, а как стратегическая проблема перераспределения когнитивных функций в человеко-машинных системах. Успешность такой интеграции определяется способностью организации сохранить примат человеческого суждения в условиях возрастающей сложности управленческих задач, обеспечивая при этом синергетический эффект от сочетания уникальных преимуществ искусственного и человеческого интеллекта.

6. ***Управление этическими, правовыми и репутационными рисками.*** Включение в процессы ИИ оценок воздействия на права человека, приватность, справедливость, внедрение мониторинга и аудита.

Внедрение искусственного интеллекта в бизнес-процессы требует формирования комплексной проактивной системы, предполагая проведение регулярных оценок воздействия алгоритмических систем на фундаментальные права человека, включая приватность и принципы справедливости.

Ключевым элементом такой системы является использование специализированного мониторинга и независимого аудита ИИ-моделей. Мониторинг позволяет отслеживать реальное поведение системы в эксплуатации, выявляя потенциальные смещения или дискриминационные последствия, которые могли не проявиться на этапе тестирования. Аудит обеспечивает верификацию соответствия ИИ-моделей не только формальным юридическим нормам, но и внутренним этическим стандартам организации, а также ожиданиям общества.

Особое значение приобретает принцип «этики по замыслу», когда этические соображения встраиваются в процесс разработки и внедрения ИИ с самого начала, а не добавляются постфактум. Это позволяет предотвратить риски

на ранних стадиях, а не бороться с их последствиями. Такой подход включает в себя оценку потенциального воздействия системы на различные группы заинтересованных лиц, анализ долгосрочных социальных последствий и создание механизмов исправления ошибок.

Управление этическими рисками ИИ — это не разовая задача, а непрерывный процесс, требующий адаптации к меняющимся социальным ожиданиям и правовым нормам. Эффективная реализация этого процесса позволяет организации не только избегать судебных разбирательств и репутационных потерь, но и укреплять доверие клиентов, сотрудников и регуляторов, что в конечном итоге создает устойчивое конкурентное преимущество.

7. ***Многоуровневая система контроля и управления:*** от технических контролей (например, тестирования безопасности) до организационных — ответственности, политик, процедур.

Создание надежной системы контроля за рисками искусственного интеллекта требует реализации многоуровневого подхода, интегрирующего технические и организационные механизмы управления. На технологическом уровне должны применяться методы активного тестирования устойчивости моделей, включая специализированные атаки на алгоритмы для выявления скрытых уязвимостей до их эксплуатации. Этот подход должен дополняться непрерывным мониторингом дрейфа данных и модельных метрик, позволяющим обнаруживать аномалии в работе систем на ранних стадиях.

Организационный уровень контроля предполагает формирование четких рамок ответственности за разработку, внедрение и эксплуатацию ИИ-систем. Создание специализированных политик и регламентов обеспечивает стандартизацию процессов управления жизненным циклом ИИ-моделей, устанавливая обязательные процедуры валидации и верификации. Внедрение системы документирования решений и архитектуры ИИ создает основу для прозрачности и подотчетности, позволяя отслеживать происхождение проблем и проводить их системный анализ.

Критически важным аспектом становится синхронизация технических и организационных контролей через создание сквозных процессов управления. Это предполагает, что результаты технического мониторинга трансформируются в управленческие решения, а организационные требования находят отражение в архитектуре и параметрах ИИ-моделей. Такой подход позволяет выстраивать динамическую систему управления рисками, способную адаптироваться к изменениям как в технологиях, так и в бизнес-среде, обеспечивая устойчивое развитие организации.

8. ***Постоянный мониторинг и адаптация.*** Поскольку ИИ-технологии развиваются быстро, необходимы регулярные ревизии, корректировки и обновления риск-профилей.

Такой мониторинг должен охватывать как поведение ИИ-систем в операционной среде, так и изменения контекста их применения. Регулярные ревизии позволяют выявлять риски, связанные с устареванием моделей, изменением

бизнес-процессов или появлением новых векторов атак. Особое значение приобретает отслеживание дрейфа данных и концепций, когда первоначально эффективная модель постепенно теряет релевантность из-за изменений в источниках информации или рыночных условиях.

Полученные в ходе мониторинга данные становятся основой для адаптации системы управления рисками. Это предполагает не только корректировку параметров конкретных моделей, но и пересмотр организационных политик, распределения ответственности и процедур контроля. Цикл «мониторинг – адаптация» замыкается через механизмы организационного обучения, когда выявленные инциденты систематически анализируются для совершенствования практик управления.

Способность к постоянному мониторингу и адаптации позволяет поддерживать соответствие системы управления рисками ИИ скорости технологических и бизнес-изменений. Это обеспечивает устойчивость компании в условиях, когда стабильность риск-ландшафта уступает место его постоянной трансформации.

Вопросы управления ИИ-рисками для менеджмента требуют внедрения KPI для управления ими. Такие показатели должны быть измеримы, понятны менеджменту и способны стимулировать результативность, связывая технические риски с бизнес-последствиями.

Для этих целей можно представить в таблице следующий структурированный набор KPI, сгруппированный по ключевым ИИ-рискам.

## Заключение

Проведенное исследование демонстрирует, что риски искусственного интеллекта для менеджмента имеют комплексную природу и затрагивают стратегические, операционные, юридические, этические и социальные аспекты деятельности организации. Исследования международных консалтинговых компаний, академических авторов и регуляторных институтов показывают, что ИИ перестал быть исключительно технологическим инструментом: он стал фактором корпоративной устойчивости, влияющим на конкурентоспособность, репутацию, распределение ответственности и структуру бизнес-процессов.

На стратегическом уровне ИИ порождает риски неправильной технической или организационной интеграции, что подтверждают примеры компаний, столкнувшихся с ошибками алгоритмов при найме, ценообразовании и сегментации клиентов. Большая часть рисков связана с данными: их качеством, корректностью маркировки, юридическим статусом и устойчивостью к сдвигам. Как показывают отраслевые опросы и аналитика, менеджеры недостаточно готовы к оценке таких рисков, а уровень алгоритмической грамотности руководителей остается низким, что создает разрыв между масштабами внедрения ИИ и компетенциями по его контролю.

Таблица

**Описание KPI менеджмента по управлению ИИ-рисками**

| Категория риска | KPI для менеджмента | Целевое значение | Что измеряет | Как измерить | Метод сбора |
|---|---|---|---|---|---|
| Стратегические риски | *Индекс стратегической объяснимости:* процент ключевых ИИ-решений, где менеджер может артикулировать логику и переопределить рекомендацию | > 90 % | Способность менеджмента понимать и оспаривать решения ИИ | (Количество ключевых бизнес-решений, по которым менеджер предоставил содержательное обоснование / Общее количество решений, принятых с помощью ИИ) × 100 % | Регистрация в системе управления решениями, опросы менеджеров |
| | *Коэффициент стратегической гибкости:* время, необходимое для переобучения/адаптации ИИ-модели при изменении рыночных условий | < [X] дней | Скорость адаптации ИИ-системы к изменениям | Среднее календарное время от выявления значительного изменения рыночных условий до переобучения и развертывания обновленной модели | Анализ логов MLOps-платформы |
| Операционные риски | *Уровень целостности данных:* процент критичных данных, соответствующих стандартам качества (полнота, точность, актуальность) | > 95 % | Надежность фундамента, на котором строятся ИИ-модели | Мониторинг метрик качества данных в режиме реального времени с помощью специализированных инструментов (например, методом Монте-Карло) | Автоматические отчеты из систем управления данными |
| | *Индекс статистической обоснованности:* процент моделей, прошедших регулярный аудит на смещение (bias) и устойчивость | 100 % | Регулярность и тщательность проверки моделей на устойчивость и смещение | (Количество моделей, прошедших плановый аудит за отчетный период / Общее количество моделей в эксплуатации) × 100 % | Отчеты отдела управления рисками моделей |
| Этические/правовые риски | *Индекс справедливости:* максимальный разрыв в метриках точности между защищенными группами | < [Y] % | Степень дискриминации моделей по отношению к защищенным группам | Расчет метрик справедливости для каждой критичной модели. KPI — максимально допустимое отклонение | Результаты регулярного тестирования моделей на смещение |

| | | | | | |
|---|---|---|---|---|---|
| | *Уровень соответствия:* процент ИИ-проектов, успешно прошедших проверку на соответствие 152-ФЗ и внутреннему этическому чек-листу | 100 % | Интеграцию этических и правовых норм в жизненный цикл ИИ | (Количество новых ИИ-проектов, одобренных Этическим комитетом и юристами / Общее количество новых ИИ-проектов) × 100 % | Данные из системы управления проектами и протоколов заседаний комитета |
| Организационные риски | *Индекс доверия персонала:* процент сотрудников, согласных с утверждением «ИИ помогает мне в работе, не ущемляя мою автономию» (из внутреннего опроса) | > 80 % | Уровень принятия ИИ сотрудниками | Проведение регулярных анонимных опросов с вопросами о восприятии ИИ, автономии и эффективности | Ежегодный или полугодовой отчет |
| | *Охват обучением:* процент менеджеров, прошедших обучение по основам ИИ и управлению ИИ-рисками | 100 % | Готовность менеджмента к работе в новой парадигме | (Количество менеджеров, завершивших обязательный курс «Управление ИИ-рисками» / Общее количество менеджеров) × 100 % | Данные из LMS (Learning Management System) |
| Финансовые риски | *Отклонение от бюджета ИИ-проектов:* разница между плановыми и фактическими затратами с учетом скрытых издержек | < [Z] % | Умение управлять скрытыми издержками (интеграция, мониторинг) | ((Фактические затраты – Плановые затраты) / Плановые затраты) × × 100 % | Данные из финансовой отчетности по проектам |
| | *Коэффициент масштабирования:* процент пилотных ИИ-проектов, дошедших до стадии промышленной эксплуатации | > 50 % | Эффективность выхода из «пилотной ловушки» | (Количество пилотов, перешедших в промышленную эксплуатацию за год / Общее количество запущенных пилотных проектов) × × 100 % | Анализ воронки развития ИИ-проектов |

*Продолжение и окончание* Таблицы

| Категория риска | KPI для менеджмента | Целевое значение | Что измеряет | Как измерить | Метод сбора |
|---|---|---|---|---|---|
| Риски для ролей | *Индекс гибридной эффективности:* рост производительности в командах, где решения принимаются человеком совместно с ИИ (против: чисто человеческие или чисто ИИ) | > [N] % | Успешность симбиоза человека и ИИ | Сравнение ключевых метрик (выручка, удовлетворенность клиентов, скорость выполнения) в командах, использующих гибридный подход, с контрольными группами | A/B тестирование и внутренний бэнч-маркетинг |
| Репутационные риски | *Индекс репутационной устойчивости:* количество негативных упоминаний в СМИ, связанных с ИИ-решениями компании | 0 | Прямое воздействие ИИ на бренд компании | Количество инцидентов, связанных с ИИ, повлекших негативные публикации в авторитетных СМИ или социальных сетях | Мониторинг медиа и социальных сетей |

Анализ показывает и то, что распространенность ИИ усиливает институциональные риски. Возникают вопросы подотчетности решений, размываются границы управленческой ответственности, а юридические требования, включая регулирование персональных данных и риск-ориентированные нормы, становятся все более сложными. Это создает правовую неопределенность и требует включения юридических рисков в систему корпоративного риск-менеджмента.

Отдельный пласт рисков имеет этический и социальный характер. Исследования демонстрируют, что непрозрачные алгоритмы могут укреплять дискриминацию, искажать социальные процессы, усиливать манипулятивные бизнес-практики и создавать асимметрию власти между компаниями и пользователями. Это подчеркивает необходимость этического аудита, внедрения механизмов объяснимости и разработки систем мониторинга последствий ИИ в реальной среде использования.

Наше исследование показывает, что существующие стандарты и регуляторные рамки создают основу для управления ИИ-рисками, но фактическая зрелость этих практик в бизнесе остается низкой. Таким образом, главные выводы нашего исследования заключаются в необходимости развития организационных компетенций, институционализации аудита ИИ, междисциплинарной координации и перехода от реактивного управления к системной интеграции рисков в корпоративную стратегию.

Особое значение наше исследование имеет для управленцев. На его основе практики-менеджеры получают понимание того, что управление рисками ИИ должно стать элементом корпоративной стратегии и не может ограничиваться техническим аудированием. Исследование предоставляет руководителям аналитическую основу, позволяющую сформировать подход к интеграции ИИ, который учитывает характер данных, характер организационных процессов, влияние на социальную среду и требования регуляторов. Для менеджеров этот текст служит ориентиром при построении систем контроля качества моделей, определении зон ответственности, распределении рисков между технологическими и бизнес-подразделениями, а также при разработке программ повышения компетентности персонала. Он показывает, что управление ИИ — это междисциплинарная задача, требующая сочетания знаний в области технологий, права, этики и корпоративного управления.

## Дискуссия

Основная дискуссионная линия исследования касается разрыва между теоретической разработанностью проблемы и ее практической реализацией в управленческих системах. Несмотря на интенсивный рост научных публикаций о безопасности ИИ и разработку множества фреймворков, компании продолжают внедрять алгоритмы быстрее, чем учатся ими управлять. В результате ИИ создает новые зоны неопределенности, которые не охвачены традиционными моделями риск-менеджмента.

Дискуссионным остается вопрос о том, насколько ИИ-риск может быть встроен в существующие корпоративные процессы. С одной стороны, сторонники классического подхода считают, что риски ИИ поддаются тем же механизмам контроля, что и другие технологические риски. С другой стороны, критики подчеркивают, что автономность, многоуровневая сложность и контекстная зависимость моделей создают принципиально новые сценарии вреда, возникающие не как технические сбои, а как изменение самой логики организационного поведения.

Дополнительное напряжение создает недостаточная алгоритмическая грамотность менеджеров. Существуют аргументы, что без понимания методов обучения, источников ошибок, вопросов смещения и объяснимости руководители не способны адекватно оценивать влияние ИИ на операционные и стратегические процессы. Необходимо поднять вопрос о необходимости включения элементов ИИ-грамотности в программы управленческого образования, что пока развито слабо.

Не менее важная часть дискуссии касается распределения ответственности. Формирующееся законодательство лишь частично освещает вопрос, кто отвечает за решения алгоритма: разработчик, интегратор, владелец модели, поставщик данных или менеджер, использующий прогнозы системы в управленческом решении. Эта неопределенность делает управление рисками ИИ нормативно уязвимым и требует создания новых институциональных механизмов подотчетности.

Наконец, существенное внимание в научных дискуссиях уделяется вопросу, как обеспечить баланс между инновациями и регуляцией. Сторонники гибких подходов предупреждают, что чрезмерный контроль может затормозить технологический прогресс. Противоположная сторона отмечает, что отсутствие строгих процедур создает социальные и экономические риски, включая массовые ошибки алгоритмов, дискриминацию и подрыв доверия к технологиям.

Наше исследование демонстрирует, что изучение рисков ИИ находится в стадии активного формирования. Несмотря на доступность множества методологических предложений, общая система управления остается незавершенной. Дальнейшее развитие области требует междисциплинарных исследований, совместной работы регуляторов, компаний и научного сообщества, а также разработки устойчивых методологических оснований для включения ИИ в корпоративные управленческие контуры.

### Список источников / References

1. Russel S. Human compatible. Artificial Intelligence and the problem of control. Viking, 2019. 336 p.

2. Tan S., Taeihagh A., Baxter K. The risks of machine learning systems // MIT Technology Review. 2022. April. P. 1–22. DOI: 10.48550/arXiv.2204.09852

3. Accountable algorithm / J. A. Kroll [et al.] // University of Pennsylvania law review. 2017. Vol. 165.3. P. 633–705. URL: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review (дата обращения: 16.11.2025).

4. Zuboff S. The age of surveillance capitalism: the fight for a human future at the new frontier of power. New York: PublicAffairs, 2019. 704 p.

5. Largest Companies View AI as a risk multiplier / D. Kingsley [et al.] // Harvard Law Scool Forum of Corporate Governance. 2024. November 20. URL: https://corpgov.law.harvard.edu/2024/11/20/largest-companies-view-ai-as-a-risk-multiplier/ (дата обращения: 18.12.2025).

### *Информация об авторе / Information about the author*

**Ольга Сергеевна Елкина** — доктор экономических наук, профессор, профессор кафедры менеджмента Северо-Западного института управления, Российская академия народного хозяйства и государственной службы при Президенте РФ, Санкт-Петербург, Россия.

**Olga Sergeevna Elkina** — Doctor of Economics, Professor, Professor of the Department of Management, North-West Institute of Management, the Russian Presidential Academy of National Economy and Public Administration, Saint Petersburg, Russia.

E-mail: phdelkina@mail.ru, https://orcid.org/0000-0003-4952-1512