*Проект*

Изображение государственного Герба Республики Казахстан

**НАЦИОНАЛЬНЫЙ СТАНДАРТ РЕСПУБЛИКИ КАЗАХСТАН**

**Информационные технологии**

**Искусственный интеллект**

**РУКОВОДСТВО ПО УПРАВЛЕНИЮ РИСКАМИ**

**СТ РК** **ISO/IEC 23894**

*(ISO/IEC 23894:2023 Information technology. Artificial intelligence. Guidance on risk management, IDT)*

*Настоящий проект стандарта не подлежит применению до его утверждения*

**Комитет технического регулирования и метрологии**

**Министерства торговли и интеграции Республики Казахстан**

**(Госстандарт)**

**Астана**

**Предисловие**

**1 ПОДГОТОВЛЕН И ВНЕСЕН** РГП на ПХВ «Казахстанский институт стандартизации и метрологии» Комитета технического регулирования и метрологии Министерства торговли и интеграции Республики Казахстан

**2 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ** Приказом Председателя Комитета технического регулирования и метрологии Министерства торговли и интеграции Республики Казахстан № \_\_ от « » \_\_\_\_ 202\_года

**3** Настоящий стандарт идентичен международному стандарту ISO/IEC 23894:2023 Information technology. Artificial intelligence. Guidance on risk management (Информационные технологии. Искусственный интеллект. Руководство по управлению рисками)

Международный стандарт ISO/IEC 23894:2023 разработан Подкомитетом SC 42 «Искусственный интеллект»

Перевод с английского языка (en)

Официальный экземпляр международного стандарта, на основе которого разработан настоящий стандарт имеется в Едином государственном фонде нормативных технических документов

Степень соответствия – идентичная (IDT)

**4 ВВЕДЕН ВПЕРВЫЕ**

*Информация об изменениях к настоящему стандарту публикуется в ежегодно издаваемом информационном каталоге «Документы по стандартизации», а текст изменений и поправок - в периодически издаваемых информационных каталогах «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в периодически издаваемом информационном каталоге «Национальные стандарты»*

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Комитета технического регулирования и метрологии Министерства торговли и интеграции Республики Казахстан.

**Содержание**

|  |  |  |
| --- | --- | --- |
| Введение | | V |
| 1 | Область применения | 1 |
| 2 | Нормативные ссылки | 1 |
| 3 | Термины и определения | 1 |
| 4 | Принципы менеджмента риска AI | 1 |
| 5 | Структура | 6 |
| 5.1 | Общие положения | 6 |
| 5.2 | Лидерство и приверженность | 7 |
| 5.3 | Интеграция | 7 |
| 5.4 | Проектирование | 7 |
| 5.5 | Реализация | 11 |
| 5.6 | Оценка | 11 |
| 5.7 | Улучшение | 11 |
| 6 | Процесс управления рисками | 11 |
| 6.1 | Общие положения | 11 |
| 6.2 | Коммуникация и консультации | 12 |
| 6.3 | Область применения, контекст и критерии | 12 |
| 6.4 | Оценка риска | 14 |
| 6.5 | Обработка риска | 17 |
| 6.6 | Мониторинг и обзор | 18 |
| 6.7 | Регистрация и отчетность | 18 |
| Приложение А *(информационное)* Цели | | 20 |
| Приложение В *(информационное)* Источники риска | | 24 |
| Приложение С *(информационное)* Менеджмент риска и жизненный цикл системы AI | | 28 |
| Библиография | | 31 |
| Приложение В.А *(информационное)* Сведения о соответствии стандартов ссылочным международным, региональным стандартам, стандартам иностранных государств | | 32 |

**Введение**

Целью менеджмента риска является создание и защита ценности. Оно повышает эффективность, поощряет инновации и поддерживает достижение целей.

Настоящий стандарт предназначен для использования в связи с ISO 31000:2018. Когда настоящий стандарт расширяет руководство, приведенное в   
ISO 31000:2018, соответствующая ссылка указывается на разделы ISO 31000:2018, за которыми следует руководство, относящееся к AI, если применимо. Для того, чтобы связь между настоящим стандартом и ISO 31000:2018 стала более явной, структура разделов ISO 31000:2018 отражена в настоящем стандарте и при необходимости внесения изменения в подразделы.

Настоящий стандарт разделен на три основные части:

Раздел 4: Принципы. В настоящем разделе описываются основополагающие принципы менеджмента риска. Использование AI требует особых рекомендаций в отношении некоторых из настоящих принципов, в соответствии с описанием в   
ISO 31000:2018 (раздел 4).

Раздел 5: Структура. Цель структуры менеджмента риска заключается в оказании помощи организации в интеграции менеджмента риска в значимые виды деятельности и функции. Аспекты, специальные для разработки, предоставления или предложения, или использования систем AI, описаны в ISO 31000:2018 (раздел 5).

Раздел 6: Процессы. Процессы менеджмента риска включают систематическое применение политик, процедур и практик к деятельности по коммуникации и консультированию, установлению контекста, а также оценке, обработке, мониторингу, обзору, регистрации и отчетности по рискам. Специализация таких процессов для AI описана в ISO 31000:2018 (раздел 6).

Общие цели и источники рисков, связанные с AI, приведены в приложении A и приложении B. В приложении C приведен пример сопоставления между процессами менеджмента риска и жизненным циклом системы AI.

**НАЦИОНАЛЬНЫЙ СТАНДАРТ РЕСПУБЛИКИ КАЗАХСТАН**

**Информационные технологии**

**Искусственный интеллект**

**РУКОВОДСТВО ПО УПРАВЛЕНИЮ РИСКАМИ**

**Дата введения \_\_\_\_ -\_\_-\_\_**

**1 Область применения**

Настоящий стандарт устанавливает рекомендации о том, как организации, которые разрабатывают, производят, внедряют или используют продукты, системы и услуги, использующие искусственный интеллект (AI), могут управлять рисками, связанными с AI. Руководство также направлено на помощь организациям в интеграции менеджмента риска в их деятельность и функции, связанные с AI. Кроме того, в нем описываются процессы для эффективного внедрения и интеграции менеджмента риска AI.

Применение настоящего стандарта может быть адаптировано к любой организации и ее контексту.

**2 Нормативные ссылки**

Для применения настоящего стандарта необходим, следующий ссылочный документ. Для датированной ссылки применяют только указанное издание ссылочного документа, для недатированной ссылки применяют последнее издание ссылочного документа (включая все его изменения):

ISO 31000:2018 Risk management. Guidelines (Менеджмент риска. Руководящие указания).

ISO Guide 73:2009 Risk management. Vocabulary (Менеджмент риска. Словарь).

ISO/IEC 22989:2022 Information technology. Artificial intelligence. Artificial intelligence concepts and terminology (Искусственный интеллект. Концепции и терминология искусственного интеллекта).

**3 Термины и определения**

В настоящем стандарте применяются термины по ISO 31000:2018,   
ISO/IEC 22989:2022 и ISO Guide 73:2009.

Примечание – ISO и IEC поддерживают терминологические базы данных для использования в стандартизации по следующим адресам:

— Платформа онлайн-просмотра ISO: доступна по адресу <https://www.iso.org/obp>.

— IEC Electropedia: доступно по адресу https://[www.electropedia.org/](http://www.electropedia.org/).

**4 Принципы менеджмента риска AI**

Менеджмент риска должен учитывать потребности организации, используя интегрированный, структурированный и всесторонний подход. Руководящие принципы позволяют организации определять приоритеты и принимать решения о том, как управлять влиянием неопределенности на ее цели.

***Проект 1 редакция***

Настоящие принципы применяются ко всем организационным уровням и целям, как стратегическим, так и операционным.

Системы и процессы обычно внедряют комбинацию различных технологий и функций в различных средах для конкретных случаев использования. Менеджмент риска должен учитывать всю систему со всеми ее технологиями и функциями, а также ее влияние на окружающую среду и заинтересованных лиц.

Системы AI могут представлять новые или возникающие риски для организации с положительными или отрицательными последствиями для целей или изменениями в вероятности существующих рисков. Они также могут потребовать специального рассмотрения со стороны организации. В настоящем документе приведены дополнительные рекомендации по принципам, структуре и процессам менеджмента риска, которые может внедрить организация.

Примечание – Различные международные стандарты имеют существенно разные определения слова «риск». В ISO 31000:2018 и связанных с ним международных стандартах «риск» подразумевает отрицательное или положительное отклонение от целей. В некоторых других международных стандартах «риск» подразумевает только потенциальные отрицательные результаты, например, проблемы, связанные с безопасностью. Настоящая разница в фокусе часто может вызывать путаницу при попытке понять и правильно внедрить соответствующий процесс менеджмента риска.

ISO 31000:2018 (раздел 4) определяет несколько общих принципов менеджмента риска. В дополнение к рекомендациям в ISO 31000:2018 (раздел 4), таблица 1 содержит дополнительные рекомендации о том, как применять такие принципы при необходимости.

**Таблица 1 - Принципы менеджмента риска, применяемые к искусственному интеллекту**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Принцип | Описание (в соответствии с указанием в  ISO 31000:2018 (раздел 4)) | Последствия для разработки и использования AI |
| a) | Интегрированный | Менеджмент риска является неотъемлемой частью всей организационной деятельности. | Нет конкретных указаний, выходящих за рамки  ISO 31000:2018. |
| b) | Структурированный и всесторонний | Структурированный и всесторонний подход к менеджменту риска способствует получению последовательных и сопоставимых результатов. | Нет конкретных указаний, выходящих за рамки  ISO 31000:2018. |
| c) | Индивидуальный | Структура и процесс менеджмента риска настраиваются и соразмерны внешнему и внутреннему контексту организации, связанному с ее целями. | Нет конкретных указаний, выходящих за рамки  ISO 31000:2018. |
| d) | Инклюзивность | Соответствующее и своевременное вовлечение заинтересованных сторон позволяет учитывать их знания, взгляды и восприятие. Это приводит к повышению осведомленности и информированному менеджменту риска. | Из-за потенциально далеко идущих последствий AI для заинтересованных сторон важно, чтобы организации стремились к диалогу с различными внутренними и внешними группами, как для сообщения о вреде и выгоде, так и для включения обратной связи и осведомленности в процесс менеджмента риска. |

*Продолжение таблицы 1*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Принцип | Описание (в соответствии с указанием в  ISO 31000:2018 (раздел 4)) | Последствия для разработки и использования AI |
|  |  |  | Организации также должны знать, что использование систем AI может привлечь дополнительные заинтересованные стороны.  Области, в которых знания, взгляды и восприятие заинтересованных сторон приносят пользу, включают, но не ограничиваются:  - Машинное обучение (МО), в частности, часто опирается на набор данных, подходящий для достижения его целей. Заинтересованные стороны могут помочь в выявлении рисков, касающихся сбора данных, операций обработки, источника и типа данных, а также использования данных для конкретных ситуаций или когда субъекты данных могут быть аномальными.  - Сложность технологий AI создает проблемы, связанные с прозрачностью и объяснимостью систем AI. Разнообразие технологий AI еще больше усугубляет настоящие проблемы из-за таких характеристик, как множественные типы модальностей данных, топологии моделей AI, а также механизмы прозрачности и отчетности, которые следует выбирать в соответствии с потребностями заинтересованных сторон. Заинтересованные стороны могут помочь определить цели и описать средства для повышения прозрачности и объяснимости систем AI. В некоторых случаях настоящие цели и средства могут быть обобщены по варианту использования и различным заинтересованным сторонам. В других случаях сегментация заинтересованными сторонами структур прозрачности и механизмов отчетности может быть адаптирована к соответствующим пользователям (например, «регуляторы», «владельцы бизнеса», «оценщики рисков модели») по варианту использования. |

*Продолжение таблицы 1*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Принцип | Описание (в соответствии с указанием в  ISO 31000:2018 (раздел 4)) | Последствия для разработки и использования AI |
|  |  |  | - Использование систем AI для автоматизированного принятия решений может напрямую влиять на внутренние и внешние заинтересованные стороны. Такие заинтересованные стороны могут предоставить свои взгляды и представления, касающиеся, например, того, где может потребоваться человеческий надзор. Заинтересованные стороны могут помочь в определении критериев справедливости, а также помочь определить, что составляет необъективность в работе системы AI. |
| e) | Динамические | Риски могут возникать, изменяться или исчезать по мере изменения внешнего и внутреннего контекста организации. Менеджмент риска предвидит, обнаруживает, признает и реагирует на настоящие изменения и события соответствующим и своевременным образом. | Для реализации руководства, предоставленного ISO 31000:2018, организации должны создать организационные структуры и меры для выявления проблем и возможностей, связанных с возникающими рисками, тенденциями, технологиями, использованием и участниками, связанными с системами AI.  Динамический менеджмент риска особенно важен для систем AI, поскольку:  - Природа систем AI сама по себе динамична из-за постоянного обучения, уточнения, оценки и проверки. Кроме того, некоторые системы AI обладают способностью адаптироваться и оптимизироваться на основе настоящего цикла, самостоятельно создавая динамические изменения.  - Ожидания клиентов в отношении систем AI высоки и могут потенциально быстро меняться, как и сами системы.  - Правовые и нормативные требования, связанные с AI, часто меняются и обновляются. |

*Продолжение таблицы 1*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Принцип | Описание (в соответствии с указанием в  ISO 31000:2018 (раздел 4)) | Последствия для разработки и использования AI |
|  |  |  | Интеграция с системами менеджмента качества, воздействия на окружающую среду, безопасности, здравоохранения, юридической или корпоративной ответственности или любой их комбинации, поддерживаемой организацией, также может рассматриваться для дальнейшего понимания и менеджмента рисками, связанными с AI, для организации, отдельных лиц и общества. |
| f) | Лучшая доступная информация | Входные данные для менеджмента риска основаны на исторической и текущей информации, а также на будущих ожиданиях. Менеджмент риска явно учитывает любые ограничения и неопределенности, связанные с такой информацией и ожиданиями. Информация должна быть своевременной, четкой и доступной для соответствующих заинтересованных сторон. | Принимая во внимание ожидание того, что AI влияет на то, как люди взаимодействуют с технологиями и реагируют на них, организациям, занимающимся разработкой систем AI, рекомендуется отслеживать соответствующую доступную информацию относительно дальнейшего использования разработанных ими систем AI, в то время как пользователи систем AI могут вести записи об использовании настоящих систем на протяжении всего срока службы системы AI.  Поскольку AI является новой технологией и постоянно развивается, статистическая информация может быть ограничена, а будущие ожидания могут быстро меняться. Организациям следует это учитывать.  Следует учитывать внутреннее использование систем AI, если таковое имеется. Отслеживание использования систем AI клиентами и внешними пользователями может быть ограничено ограничениями в области интеллектуальной собственности, договорными или рыночными ограничениями. Такие ограничения должны быть зафиксированы в процессе менеджмента риска AI и обновлены, когда условия ведения бизнеса требуют пересмотра. |

*Продолжение таблицы 1*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Принцип | Описание (в соответствии с указанием в  ISO 31000:2018 (раздел 4)) | Последствия для разработки и использования AI |
| g) | Человеческие и культурные факторы | Человеческое поведение и культура существенно влияют на все аспекты менеджмента риска на каждом уровне и этапе. | Организации, занимающиеся проектированием, разработкой или внедрением систем AI или любой их комбинацией, должны контролировать человеческий и культурный ландшафт (среду), в котором они находятся. Организациям следует сосредоточиться на определении того, как системы или компоненты AI взаимодействуют с уже существующими социальными моделями, которые могут привести к воздействию на справедливые результаты, конфиденциальность, свободу выражения мнений, справедливость, безопасность, занятость, окружающую среду и права человека в целом. |
| h) | Постоянное улучшение | Менеджмент риска постоянно улучшается посредством обучения и опыта. | Выявление ранее неизвестных рисков, связанных с использованием систем AI, следует учитывать в процессе постоянного улучшения. Организации, занимающиеся проектированием, разработкой или внедрением систем AI или компонентов систем или любой их комбинацией, должны отслеживать экосистему AI на предмет успехов в производительности, недостатков и извлеченных уроков, а также поддерживать осведомленность о новых результатах и методах исследований AI (возможностях для улучшения). |

**5 Структура**

**5.1 Общие положения**

Цель структуры менеджмента риска - помочь организации интегрировать менеджмент риска в значимые виды деятельности и функции. Применяется руководство, представленное в ISO 31000:2018, 5.1.

Менеджмент риска включает сбор соответствующей информации для организации для принятия решений и устранения рисков. В то время как руководящий орган определяет общую готовность к риску и организационные цели, он делегирует процесс принятия решений по выявлению, оценке и обработке рисков руководству внутри организации.

[1] описывает дополнительные рекомендации по управлению для организации в отношении разработки, покупки или использования системы AI. Такие рекомендации включают новые возможности, потенциальные изменения в готовности к риску, а также новые политики управления для обеспечения ответственного использования AI организацией. Его можно использовать в сочетании с процессами менеджмента риска, описанными в настоящем документе, чтобы помочь направить динамическую и итеративную (повторяющуюся) организационную интеграцию, описанную в   
ISO 31000:2018, 5.2.

**5.2 Лидерство и приверженность**

Применяются рекомендации, представленные в ISO 31000:2018, 5.2.

В дополнение к рекомендациям, представленным в ISO 31000:2018, 5.2, применяется следующее:

В связи с особой важностью доверия и подотчетности, связанных с разработкой и использованием AI, высшее руководство должно рассмотреть вопрос о том, как политики и заявления, связанные с рисками AI и управлением рисками, доводятся до сведения заинтересованных сторон. Демонстрация такого уровня лидерства и приверженности может иметь решающее значение для обеспечения уверенности заинтересованных сторон в том, что AI разрабатывается и используется ответственно.

Поэтому организация должна рассмотреть возможность выпуска заявлений, связанных с ее приверженностью менеджменту риска AI, чтобы повысить уверенность заинтересованных сторон в использовании ими AI.

Высшее руководство также должно знать о специализированных ресурсах, которые могут потребоваться для менеджмента риска AI, и распределять настоящие ресурсы соответствующим образом.

**5.3 Интеграция**

Применяется руководство, представленное в ISO 31000:2018, 5.3.

**5.4 Проектирование**

5.4.1 Понимание организации и ее контекста

Применяется руководство, представленное в ISO 31000:2018, 5.4.1.

В дополнение к рекомендациям, представленным в ISO 31000:2018, 5.4.1, в   
таблице 2 перечислены дополнительные факторы, которые следует учитывать при понимании внешнего контекста организации.

**Таблица 2 - Рассмотрение при установлении внешнего контекста организации**

|  |  |
| --- | --- |
| **Общие указания, представленные в ISO 31000:2018, 5.4.1**  Организации должны учитывать, как минимум следующие элементы своего внешнего контекста: | **Дополнительные указания для организаций, занимающихся AI**  Организации должны дополнительно учитывать, но не исключительно, следующие элементы: |
| - Социальные, культурные, политические, правовые, нормативные, финансовые, технологические, экономические и экологические факторы, будь то международные, национальные, | - Соответствующие правовые требования, включая те, которые конкретно относятся к AI.  - Руководства по этическому использованию и проектированию AI и автоматизированных систем, выпущенные государственными |

*Продолжение таблицы 2*

|  |  |
| --- | --- |
| **Общие указания, представленные в ISO 31000:2018, 5.4.1**  Организации должны учитывать, как минимум следующие элементы своего внешнего контекста: | **Дополнительные указания для организаций, занимающихся AI**  Организации должны дополнительно учитывать, но не исключительно, следующие элементы: |
| региональные или локальные; | группами, регулирующими органами, органами стандартизации, гражданским обществом, академическими кругами и отраслевыми ассоциациями.  - Конкретныее для предметной области руководства и структуры, связанные с AI. |
| - Ключевые движущие силы и тенденции, влияющие на цели организации; | - Технологические тенденции и достижения в различных областях AI.  - Социальные и политические последствия внедрения систем AI, включая рекомендации социальных наук. |
| - Отношения, восприятие, ценности, потребности и ожидания внешних заинтересованных сторон; | - Восприятие заинтересованных сторон, на которое могут влиять такие проблемы, как отсутствие прозрачности (также называемое непрозрачностью) систем AI или необъективность систем AI.  - Ожидания заинтересованных сторон относительно доступности конкретных решений на основе AI и средств, с помощью которых модели AI становятся доступными (например, через пользовательский интерфейс, комплект для разработки программного обеспечения). |
| - Договорные отношения и обязательства; | - Как использование AI, особенно систем AI, использующих постоянное обучение, может повлиять на способность организации выполнять договорные обязательства и гарантии. Следовательно, организациям следует тщательно рассмотреть объем соответствующих договоров.  - Договорные отношения при проектировании и производстве систем и услуг AI. Например, права собственности и использования тестовых и обучающих данных должны учитываться при предоставлении их третьими лицами |
| - Сложность сетей и зависимостей; | - Использование AI может повысить сложность сетей и зависимостей. |
| - (руководство кроме ISO 31000:2018). | - Система AI может заменить существующую систему, и в таком случае можно провести оценку преимуществ и рисков передачи системы AI по сравнению с существующей системой, учитывая вопросы безопасности, окружающей среды, социальные, технические и финансовые вопросы, связанные с внедрением системы AI |

В дополнение к рекомендациям, изложенным в ISO 31000:2018, 5.4.1, в таблице 3 перечислены дополнительные факторы, которые следует учитывать при понимании внутреннего контекста организации.

**Таблица 3 - Рассмотрение при установлении внутреннего контекста организации**

|  |  |
| --- | --- |
| **Общее руководство, предоставленное ISO 31000:2018, 5.4.1**  Организации должны учитывать, как минимум следующие элементы своего внутреннего контекста: | **Дополнительное руководство для организаций, занимающихся AI**  Организации должны дополнительно учитывать, но не исключительно, следующие элементы |
| - Видение, миссия и ценности; | - Никаких конкретных указаний, помимо  ISO 31000:2018 |
| - Управление, организационная структура, роли и подотчетность; | - Никаких конкретных указаний, помимо  ISO 31000:2018 |
| - Стратегия, цели и политика; | - Никаких конкретных указаний, помимо  ISO 31000:2018 |
| - Культура организации; | - Влияние, которое система AI может оказать на культуру организации, изменяя и вводя новые обязанности, роли и задачи. |
| - Стандарты, руководящие принципы и модели, принятые организацией; | - Любые дополнительные международные, региональные, национальные и местные стандарты и руководящие принципы, которые налагаются при использовании систем AI. |
| - Возможности, понимаемые с точки зрения ресурсов и знаний (например, капитал, время, люди, интеллектуальная собственность, процессы, системы и технологии); | - Дополнительные риски для организационных знаний, связанные с прозрачностью и объяснимостью систем AI.  - Использование систем AI может привести к изменению количества человеческих ресурсов, необходимых для реализации определенной возможности, или к изменению типа необходимых ресурсов, например, к деквалификации или потере опыта, когда принятие решений человеком все больше поддерживается системами AI.  - Конкретные знания в области технологий AI и науки о данных, необходимые для разработки и использования систем AI.  - Доступность инструментов, платформ и библиотек AI может позволить разрабатывать системы AI без полного понимания технологии, ее ограничений и потенциальных подводных камней.  - Потенциал AI в плане возникновения проблем и возможностей, связанных с интеллектуальной собственностью для конкретных систем AI. Организации должны рассмотреть свою собственную интеллектуальную собственность в настоящей области и то, как интеллектуальная собственность может повлиять на прозрачность, безопасность и способность сотрудничать с заинтересованными сторонами, чтобы определить, следует ли предпринимать какие-либо шаги. |

*Продолжение таблицы 3*

|  |  |
| --- | --- |
| **Общее руководство, предоставленное  ISO 31000:2018, 5.4.1**  Организации должны учитывать, как минимум следующие элементы своего внутреннего контекста: | **Дополнительное руководство для организаций, занимающихся AI**  Организации должны дополнительно учитывать, но не исключительно, следующие элементы |
| - Данные, информационные системы и информационные потоки; | - Системы AI могут использоваться для автоматизации, оптимизации и улучшения обработки данных.  - Как потребители данных, системы AI могут накладывать дополнительные ограничения на качество и полноту данных и информации. |
| - Отношения с внутренними заинтересованными сторонами, принимая во внимание их восприятие и ценности; | - Восприятие заинтересованных сторон, на которое могут влиять такие проблемы, как отсутствие прозрачности систем AI или необъективность систем AI.  - Потребности и ожидания заинтересованных сторон могут быть удовлетворены в большей степени конкретными системами AI.  - Необходимость информирования заинтересованных сторон о возможностях, режимах отказов и управлении отказами систем AI |
| - Договорные отношения и обязательства; | - Восприятие заинтересованных сторон, на которое могут влиять различные проблемы, связанные с системами AI, такие как потенциальное отсутствие прозрачности и несправедливость.  - Потребности и ожидания заинтересованных сторон могут быть удовлетворены определенными системами AI.  - Необходимость информирования заинтересованных сторон о возможностях, режимах отказов и управлении отказами систем AI.  - Ожидания заинтересованных сторон относительно конфиденциальности, а также индивидуальных и коллективных основных прав, и свобод. |
| - Взаимозависимости и взаимосвязи; | - Использование систем AI может повысить сложность взаимозависимостей и взаимосвязей. |

В дополнение к руководству, представленному в ISO 31000:2018, 5.4.1, организации должны учитывать, что использование систем AI может увеличить потребность в специализированном обучении.

5.4.2 Формулирование обязательств по менеджменту риска

Применяется руководство, представленное в ISO 31000:2018, 5.4.2.

5.4.3 Назначение организационных ролей, полномочий, ответственности и подотчетности

Применяется руководство, представленное в ISO 31000:2018, 5.4.3.

В дополнение к руководству ISO 31000:2018, 5.4.3, высшее руководство и надзорные органы, где это применимо, должны выделять ресурсы и определять лиц:

- с полномочиями по устранению рисков AI;

- с ответственностью за установление и мониторинг процессов для устранения рисков AI.

5.4.4 Распределение ресурсов

Применяется руководство, представленное в ISO 31000:2018, 5.4.4.

5.4.5 Установление коммуникации и консультаций

Применяется руководство, изложенное в ISO 31000:2018, 5.4.5.

**5.5 Реализация**

Применяется руководство, представленное в ISO 31000:2018, 5.5.

**5.6 Оценка**

Применяется руководство, представленное в ISO 31000:2018, 5.6.

**5.7 Улучшение**

5.7.1 Адаптация

Применяется руководство, представленное в ISO 31000:2018, 5.7.1.

5.7.2 Постоянное улучшение

Применяется руководство, представленное в ISO 31000:2018, 5.7.2.

**6 Процесс менеджмента риска**

**6.1 Общие положения**

Применяется руководство, представленное в ISO 31000:2018, 6.1.

Организации должны внедрять подход, основанный на оценке рисков, для выявления, оценки и понимания рисков AI, которым они подвергаются, и принимать соответствующие меры по их обработке в соответствии с уровнем риска. Успех общего процесса менеджмента риска AI в организации зависит от идентификации, создания и успешной реализации узконаправленных процессов менеджмента риска на стратегическом, операционном, программном и проектном уровнях. Из-за проблем, связанных, помимо прочего, с потенциальной сложностью, отсутствием прозрачности и непредсказуемостью некоторых технологий на основе AI, особое внимание следует уделять процессам менеджмента риска на уровне проекта системы AI. Настоящие процессы на уровне проекта системы должны быть согласованы с целями организации и должны быть информированы и информировать другие уровни менеджмента риска. Например, эскалации и уроки, извлеченные на уровне проекта AI, должны быть включены на более высоких уровнях, таких как стратегический, операционный и программный уровни, и другие, если применимо.

Область применения, контекст и критерии процесса менеджмента риска на уровне проекта напрямую зависят от этапов жизненного цикла системы AI, которые входят в область применения проекта. В приложении C показаны возможные связи между процессом менеджмента риска на уровне проекта и жизненным циклом системы AI (в соответствии с определением в ISO/IEC 22989:2022).

**6.2 Коммуникация и консультации**

Применяется руководство, представленное в ISO 31000:2018, 6.2.

Список заинтересованных сторон, на которых могут повлиять системы AI, может быть больше, чем предполагалось изначально, может включать в себя иным образом неучтенные внешние заинтересованные стороны и может распространяться на другие части общества.

**6.3 Область применения, контекст и критерии**

6.3.1 Общие положения

Применяется руководство, представленное в ISO 31000:2018, 6.3.1.

В дополнение к рекомендациям, представленным в ISO 31000:2018, 6.3.1, для организаций, использующих AI, следует расширить область менеджмента риска AI, контекст процесса менеджмента риска AI и критерии оценки значимости риска для поддержки процессов принятия решений, чтобы определить, где системы AI разрабатываются или используются в организации. Такой перечень разработки и использования AI должен быть задокументирован и включен в процесс менеджмента риска организации.

6.3.2 Определение области применения

Применяется руководство, представленное в ISO 31000:2018, 6.3.2.

Область применения должна учитывать конкретные задачи и обязанности различных уровней организации. Кроме того, следует учитывать цели и назначение систем AI, разработанных или используемых организацией.

6.3.3 Внешний и внутренний контекст

Применяется руководство, представленное в ISO 31000:2018, 6.3.3.

Из-за масштабов потенциального воздействия систем AI организация должна уделять особое внимание среде своих заинтересованных сторон при формировании и установлении контекста процесса менеджмента риска.

Следует уделить внимание рассмотрению списка заинтересованных сторон, включая, помимо прочего:

- организацию (саму);

- клиентов, партнеров и сторонних лиц;

- поставщиков;

- конечных пользователей;

- регулирующие органы;

- гражданские организации;

- физических лиц;

- затронутые сообщества;

- общества.

Вот некоторые другие рекомендации для внешнего и внутреннего контекста:

- могут ли системы AI нанести вред людям, лишить основных услуг (которые, если их прервать, поставят под угрозу жизнь, здоровье или личную безопасность) или нарушить права человека (например, из-за несправедливого и необъективного автоматизированного принятия решений) или способствовать нанесению вреда окружающей среде;

- внешние и внутренние ожидания в отношении социальной ответственности организации;

- внешние и внутренние ожидания в отношении экологической ответственности организации.

Руководящие указания в ISO 26000:2010[2], описывающие аспекты социальной ответственности, должны применяться в качестве основы для понимания и обработки рисков, особенно в отношении основных тем организационного управления, прав человека, трудовой практики, окружающей среды, добросовестной деловой практики, проблем потребителей, а также вовлечения и развития сообщества.

Примечание - Дополнительная справочная информация о надежности доступна в [3].

6.3.4 Определение критериев риска

Применяется руководство, представленное в ISO 31000:2018, 6.3.4.

В дополнение к рекомендациям, изложенным в ISO 31000:2018, 6.3.4, в таблице 4 приведены дополнительные рекомендации по факторам, которые следует учитывать при определении критериев риска:

**Таблица 4 - Дополнительные рекомендации при определении критериев риска**

|  |  |
| --- | --- |
| Рекомендации по определению критериев риска, в соответствии с указанием в  ISO 31000:2018, 6.3.4 | Дополнительные рекомендации в контексте разработки и использования систем AI |
| - Характер и тип неопределенностей, которые могут повлиять на результаты и цели (как материальные, так и нематериальные); | - Организации должны предпринять обоснованные шаги для понимания неопределенности во всех частях системы AI, включая используемые данные, программное обеспечение, математические модели, физическое расширение и аспекты человеческого участия в системе (например, любая связанная человеческая деятельность во время сбора и маркировки данных). |
| - Как будут определяться и измеряться последствия (как положительные, так и отрицательные) и их вероятность; |
| - Факторы, связанные со временем; | - Нет конкретных указаний, выходящих за рамки ISO 31000:2018 |
| - Последовательность в использовании измерений; | - Организации должны знать, что AI - это быстро развивающаяся технологическая область. Методы измерения должны последовательно оцениваться в соответствии с их эффективностью и пригодностью для используемых систем AI. |
| - Как определить уровень риска; | - Организации должны установить последовательный подход к определению уровня риска. Подход должен отражать потенциальное влияние систем AI на различные цели, связанные с AI (см. Приложение A). |
| - Как будут учитываться комбинации и последовательности множественных рисков; | - Никаких конкретных указаний, выходящих за рамки ISO 31000:2018 |
| - Возможности организации. | - Возможности организации в области AI, уровень знаний и способность смягчать реализованные риски AI следует учитывать при принятии решения о ее готовности к риску в области AI. |

**6.4 Оценка рисков**

6.4.1 Общие положения

Применяется руководство, представленное в ISO 31000:2018, 6.4.1.

Риски AI должны быть идентифицированы, количественно оценены или качественно описаны и ранжированы по приоритетам в соответствии с критериями риска и целями, относящимися к организации. В приложении B представлен примерный каталог источников риска, связанных с AI. Такой примерный каталог не может считаться исчерпывающим. Тем не менее, опыт показал ценность использования такого каталога в качестве основы для любой организации, впервые выполняющей оценку рисков или интегрирующей менеджмент риска AI в существующие структуры управления. Каталог служит документированной основой для настоящих организаций.

Поэтому организации, занимающиеся разработкой, предоставлением или применением систем AI, должны согласовывать свою деятельность по оценке рисков с жизненным циклом системы. Различные методы оценки рисков могут применяться к различным этапам жизненного цикла системы.

6.4.2 Идентификация риска

6.4.2.1 Общие положения

Применяется руководство, представленное в ISO 31000:2018, 6.4.2.

6.4.2.2 Определение активов и их ценности

Организация должна определить активы, связанные с проектированием и использованием AI, которые попадают в область процесса менеджмента риска, в соответствии с определением в 6.3.2. Понимание того, какие активы находятся в области применения, а также относительной критичности или ценности настоящих активов является неотъемлемой частью оценки воздействия. Следует учитывать как стоимость актива, так и его характер (материальный или нематериальный). Кроме того, в отношении разработки и использования AI, активы следует рассматривать в контексте элементов, включая, но не ограничиваясь следующим:

- Активы и их ценность для организации:

- Материальные активы могут включать данные, модели и саму систему AI.

- Нематериальные активы могут включать репутацию и доверие.

- Активы и их ценность для отдельных лиц:

- Материальные активы могут включать персональные данные отдельного лица,

- Нематериальные активы могут включать конфиденциальность, здоровье и безопасность отдельного лица.

- Активы и их ценность для сообществ и обществ:

- Материальные активы могут включать окружающую среду,

- Нематериальные активы, вероятно, больше основаны на ценностях, таких как социокультурные убеждения, знания сообщества, доступ к образованию и справедливость.

Оценки активов и связи с воздействием см. 6.4.2.6 и 6.4.3.2.

Примечание - Использование слова «актив» с иллюстративными примерами в настоящем разделе не имеет никаких юридических последствий.

6.4.2.3 Определение источников риска

Организация должна определить список источников риска, связанных с разработкой или использованием AI, или и тем, и другим, в рамках определенной области.

Источники риска могут быть определены в следующих областях, но не ограничиваться ими:

- организация;

- процессы и процедуры;

- процедуры управления;

- персонал;

- физическая среда;

- данные;

- конфигурация системы AI;

- среда внедрения;

- оборудование, программное обеспечение, сетевые ресурсы и услуги;

- зависимость от внешних сторон.

Примеры источников риска, связанных с AI, можно найти в Приложении B.

6.4.2.4 Определение потенциальных событий и результатов

Организация должна определить потенциальные события, которые связаны с разработкой или использованием AI и могут привести к различным материальным или нематериальным последствиям.

События могут быть определены с помощью одного или нескольких из следующих методов и источников:

- опубликованные стандарты;

- опубликованные технические спецификации;

- опубликованные технические отчеты;

- опубликованные научные статьи;

- рыночные данные по аналогичным системам или приложениям, которые уже используются;

- отчеты об инцидентах в аналогичных системах или приложениях, которые уже используются;

- полевые испытания;

- исследования удобства использования;

- результаты соответствующих исследований;

- отчеты заинтересованных сторон;

- собеседования с внутренними или внешними экспертами и отчеты от них;

- моделирование.

6.4.2.5 Определение средств контроля

Организация должна определить средства контроля, имеющие отношение либо к разработке, либо к использованию AI, либо к тому и другому. Средства контроля должны быть определены в ходе мероприятий по менеджменту риска и задокументированы (во внутренних системах, процедурах, аудиторских отчетах и т. д.).

Средства контроля могут использоваться для положительного влияния на общий риск путем смягчения источников риска, событий и результатов.

Также следует учитывать операционную эффективность определенных средств контроля, особенно сбои в контроле.

6.4.2.6 Определение последствий

В рамках оценки рисков AI, организация должна определить источники риска, события или результаты, которые могут привести к рискам. Она также должна определить любые последствия для самой организации, для отдельных лиц, сообществ, групп и обществ. Организации должны проявлять особую осторожность, чтобы выявить любые различия между группами, которые испытывают преимущества технологии, и группами, которые испытывают негативные последствия.

Последствия для организации обязательно отличаются от последствий для физических лиц и обществ. Последствия для организаций могут включать, но не ограничиваются:

- время расследования и ремонта;

- (рабочее) время, полученное и потерянное;

- возможности, полученные или потерянные;

- угрозы здоровью или безопасности отдельных лиц;

- финансовые затраты на определенные навыки для устранения ущерба;

- набор, выполнение требований и удержание сотрудников;

- имидж, репутация и доброжелательность;

- штрафы и пени;

- судебные разбирательства с клиентами.

В зависимости от контекста последствия для отдельных лиц и обществ могут быть более общими, и в настоящем случае организация может быть не в состоянии точно оценить, каков эффект для каждого отдельного лица или для обществ.

Вместо того, чтобы указывать каждый тип эффекта, его можно рассматривать в целом, поскольку ключевым элементом является степень критичности эффектов (например, для конфиденциальности, справедливости, прав человека и т. д. в случае отдельного лица или для окружающей среды в случае обществ).

Точные эффекты могут зависеть от контекста, в котором работает организация, и областей, для которых разрабатывается и используется система AI.

Примечания

1 Последствия могут быть положительными или отрицательными. Организация может учитывать оба варианта при оценке последствий для организации, для отдельных лиц и для обществ.

2 Последствия для отдельных лиц и обществ обычно также могут приводить к последствиям для организации. Например, инцидент безопасности для пользователя продукта организации может привести к искам об ответственности к организации и отрицательно повлиять на ее репутацию и продажи продукции.

6.4.3 Анализ риска

6.4.3.1 Общие положения

Применяется руководство, представленное в ISO 31000:2018, 6.4.3.

Подход к анализу должен соответствовать критериям риска, разработанным в рамках установления контекста (см. 6.3).

6.4.3.2 Оценка последствий

При оценке последствий, выявленных в оценке рисков, организация должна различать оценку воздействия на бизнес, оценку воздействия на отдельных лиц и оценку воздействия на общество.

Анализы воздействия на бизнес должны определять степень воздействия на организацию и учитывать элементы, включая, но не ограничиваясь следующим:

- критичность воздействия;

- ощутимые и нематериальные воздействия;

- критерии, используемые для установления общего воздействия (в соответствии с определением в 6.3.4).

Анализы воздействия на отдельных лиц должны определять степень, в которой на отдельного человека может повлиять разработка или использование AI организацией, или и то, и другое. Они должны учитывать элементы, включая, но не ограничиваясь следующим:

- типы данных, используемых от отдельных лиц;

- предполагаемое воздействие разработки или использования AI;

- потенциальное воздействие необъективности на человека;

- потенциальное воздействие на основные права, которое может привести к материальному и нематериальному ущербу для человека;

- потенциальное воздействие на справедливость для человека;

- безопасность человека;

- защита и смягчающие меры контроля вокруг нежелательной необъективности и несправедливости;

- юрисдикционная и культурная среда отдельного человека (которая может влиять на то, как определяется относительное воздействие).

Анализ воздействия для обществ должен определять степень, в которой общества могут быть затронуты либо разработкой, либо использованием AI организацией, либо и тем, и другим. Они должны учитывать элементы, включая, но не ограничиваясь следующим:

- область применения общественного воздействия (насколько широк охват системы AI в различных группах населения), включая то, кем система используется или для кого она разработана (например, государственное использование может потенциально влиять на общества больше, чем частное использование);

- как система AI влияет на социальные и культурные ценности, которых придерживаются различные затронутые группы (включая конкретные способы, которыми система AI усиливает или уменьшает уже существующие модели вреда для различных социальных групп).

6.4.3.3 Оценка вероятности

Если применимо, организация должна оценивать вероятность возникновения событий и результатов, вызывающих риски. Вероятность может быть определена по качественной или количественной шкале и должна соответствовать критериям, установленным в рамках 6.3.4. Вероятность может быть информирована и на нее могут влиять (не ограничиваясь):

- типы, значимость и количество источников риска;

- частота, серьезность и распространенность угроз;

- внутренние факторы, такие как операционный успех политик и процедур и мотивация внутренних субъектов;

- внешние факторы, такие как география и другие социальные, экономические и экологические проблемы;

- успех (смягчение) или неудача средств контроля (см. 6.4.2.5).

Организации должны включать расчеты вероятности только там, где они применимы и полезны для определения того, где применять обработку рисков. Могут возникнуть значительные технические, экономические и эвристические проблемы с вероятностями принятия решений, особенно когда вероятность либо невозможно рассчитать, либо когда расчет имеет большую погрешность.

6.4.4 Оценка риска

Применяется руководство, приведенное в ISO 31000:2018, 6.4.4.

**6.5 Обработка риска**

6.5.1 Общие положения

Применяется руководство, приведенное в ISO 31000:2018, 6.5.1.

6.5.2 Выбор вариантов обработки риска

Применяется руководство, представленное в ISO 31000:2018, 6.5.2.

Определенные организацией варианты обработки риска должны быть разработаны для снижения негативных последствий рисков до приемлемого уровня и повышения вероятности достижения положительных результатов. Если требуемое снижение негативных результатов не может быть достигнуто путем применения различных вариантов обработки риска, организация должна провести анализ риска и выгоды для остаточных рисков.

В соответствии с ISO 31000:2018, 6.5.2 организация должна рассмотреть:

- не допущение риска путем принятия решения не начинать или не продолжать деятельность, которая приводит к риску;

- принятие или увеличение риска для реализации возможности;

- устранение источника риска;

- изменение вероятности;

- изменение последствий;

- совместное принятие риска (например, посредством контрактов или покупки страховки);

- удержание риска посредством обоснованного решения.

6.5.3 Подготовка и реализация планов обработки риска

Применяется руководство, приведенное в ISO 31000:2018, 6.5.3.

После того, как план обработки риска будет задокументирован, следует внедрить меры обработки риска, выбранные в 6.5.2.

Внедрение каждой меры обработки риска и ее эффективность должны быть проверены и зафиксированы в соответствии с 6.7.

**6.6 Мониторинг и обзор**

Применяется руководство, приведенное в ISO 31000:2018, 6.6.

**6.7 Регистрация и отчетность**

Применяется руководство, приведенное в ISO 31000:2018, 6.7.

Организация должна создать, зарегистрировать и поддерживать систему сбора и верификации информации о продукте или аналогичных продуктах с этапов реализации и после реализации. Организация также должна собирать и просматривать общедоступную информацию о подобных системах на рынке.

Затем настоящая информация должна быть оценена на предмет возможной релевантности для надежности системы AI. В частности, оценка должна оценить, существуют ли ранее необнаруженные риски или ранее оцененные риски больше не являются приемлемыми. Настоящая информация может быть введена и учтена в процессе менеджмента риска AI организации в качестве корректировки целей, вариантов использования или извлеченных уроков.

Если применяется какое-либо из настоящих условий, организации должны выполнить следующее:

- оценить влияние на предыдущие действия по менеджменту риска и вернуть результаты настоящей оценки в процесс менеджмента риска.

- провести обзор действий по менеджменту риска для системы AI. Если есть вероятность, что остаточный риск или его принятие изменились, следует оценить влияние на существующие меры контроля рисков.

Результаты настоящей оценки должны быть зарегистрированы. Запись менеджмента риска должна позволять отслеживать каждый выявленный риск во всех процессах менеджмента риска. Записи могут использовать общий шаблон, согласованный организацией.

В дополнение к документации области применения, контекста и критериев (см. 6.3), оценки риска (см. 6.4) и обработки риска (см. 6.5), запись должна включать как минимум следующую информацию:

- описание и идентификацию системы, которая была проанализирована;

- примененная методика;

- описание предполагаемого использования системы AI;

- идентификация лица(лиц) и организации, которая провела оценку риска;

- поставленные задачи и дата оценки риска;

- статус выпуска оценки риска;

- были ли достигнуты цели и в какой степени.

**Приложение А**

*(информационное)*

**Цели**

**A.1 Общие положения**

При определении рисков систем AI следует учитывать различные цели, связанные с AI, в зависимости от характера рассматриваемой системы и контекста ее применения. Цели, связанные с AI, которые следует учитывать, включают, помимо прочего, цели, описанные в разделах A.2 по A.12.

**A.2 Подотчетность**

Подотчетность относится как к характеристике организаций, так и к системному свойству:

- Организационная подотчетность означает, что организация берет на себя ответственность за свои решения и действия, объясняя их и отвечая за них перед руководящим органом, юридическими органами и, в более широком смысле, перед заинтересованными сторонами.

- Системная подотчетность относится к возможности отслеживать решения и действия субъекта к настоящему субъекту.

Использование AI может изменить существующие структуры подотчетности. Если ранее лица выполняли действия, за которые они несли бы ответственность, то теперь такие действия могут полностью или частично выполняться системами AI. Кто будет нести ответственность в настоящем случае, является постоянным рассмотрением регулирующих органов. Разработчики и пользователи систем AI должны знать соответствующее законодательство в странах, где система выводится на рынок и используется.

**A.3 Экспертиза в области AI**

Системы AI и их разработка отличаются от программных решений, не связанных с AI. Необходим выбор специализированных специалистов с междисциплинарными навыками и опытом в оценке, разработке и внедрении систем AI. Организации должны гарантировать, что люди с такими знаниями участвуют в разработке и спецификации систем AI.

Экспертиза в области AI должна распространяться на конечных пользователей систем AI. Пользователи должны иметь достаточное понимание того, как функционирует система AI, и иметь возможность обнаруживать и отменять ошибочные решения или результаты.

**A.4 Доступность и качество данных для обучения и тестирования**

Системам AI на основе МО требуются данные для обучения и тестирования, чтобы обучить и проверить системы на предмет предполагаемого поведения. Развернутая система AI работает на производственных данных. Данные для обучения, тестирования и производства должны соответствовать предполагаемому поведению с точки зрения типа данных и качества.

Данные для обучения и тестирования должны быть проверены на предмет их актуальности и важности для предполагаемой цели. Объем требуемых данных для обучения и тестирования может варьироваться в зависимости от предполагаемой функциональности и сложности среды. Данные для обучения и тестирования должны иметь достаточно разнообразные характеристики, чтобы обеспечить высокую предсказательную силу для системы AI. Кроме того, необходимо обеспечить согласованность данных для обучения и тестирования, используя при этом независимые наборы данных, когда это применимо.

Возможно, данные для обучения и тестирования недоступны в компании и получены извне. В настоящем случае также необходимо обеспечить качество данных.

**A.5 Воздействие на окружающую среду**

Использование AI может вызывать эффекты с точки зрения экологии. Использование AI может оказывать положительное воздействие на окружающую среду. Например, система AI может использоваться для снижения содержания оксида азота в газовой турбине. Использование AI также может оказывать отрицательное воздействие на окружающую среду из-за интенсивного использования ресурсов. Например, этап обучения некоторых систем AI требует вычислительных ресурсов и может потреблять значительное количество электроэнергии. Настоящие воздействия на окружающую среду следует учитывать.

**A.6 Справедливость**

Использование систем AI для автоматизированного принятия решений может быть несправедливым по отношению к определенным лицам или группам лиц. Несправедливые результаты имеют ряд причин, такие как необъективность в целевых функциях, несбалансированные наборы данных и человеческая необъективность в обучающих данных и в предоставлении обратной связи системам. Несправедливость также может быть вызвана проблемой необъективности в концепции продукта, формулировке проблемы или выборе того, когда и где развертывать системы AI.

Для получения дополнительной информации о необъективности и справедливости в системах AI см. [4].

**A.7 Техническая поддерживаемость**

Поддерживаемость связана со способностью организации обрабатывать модификации системы AI для исправления дефектов или адаптации к новым требованиям. Поскольку системы AI, основанные на МО, обучаются и не следуют подходу, основанному на правилах, следует изучить поддерживаемость системы AI и ее последствия.

**A.8 Конфиденциальность**

Конфиденциальность связана со способностью отдельных лиц контролировать или влиять на то, какая информация, связанная с ними, может быть собрана, сохранена и обработана, и кем настоящая информация может быть раскрыта. Как поясняется в [3], «многие методы AI (например, глубокое обучение) в значительной степени зависят от больших данных, поскольку их точность зависит от объема используемых ими данных. Неправильное использование или раскрытие некоторых данных, особенно личных и конфиденциальных данных (например, медицинских карт), может иметь пагубные последствия для субъектов данных. Таким образом, защита конфиденциальности стала серьезной проблемой в анализе больших данных и AI».

Следует рассмотреть возможность определения того, может ли система AI выводить конфиденциальные персональные данные. Для систем AI защита конфиденциальности включает защиту данных, используемых для создания и эксплуатации системы AI, обеспечение того, чтобы система AI не могла использоваться для предоставления необоснованного доступа к ее данным, и защиту доступа к моделям, персонализированным для отдельного человека или которые могут использоваться для вывода информации или характеристик похожих людей.

Неправильный сбор, использование и раскрытие личной информации также могут иметь прямое влияние на основные права человека, такие как дискриминация и свобода выражения мнений и информации. Также следует учитывать влияние на этические принципы с точки зрения уважения человеческих ценностей и человеческого достоинства.

Примечание - Оценка воздействия на защиту данных (см. [5]), часто называемая оценкой воздействия на конфиденциальность, является полезным инструментом для менеджмента риска, связанными с использованием персональных данных во время сбора данных, обучения системы AI и использования системы AI.

**A.9 Надежность**

Надежность связана со способностью системы поддерживать свой уровень производительности в различных обстоятельствах ее использования. Следует учитывать степень, в которой система AI или связанный с ней компонент могут правильно функционировать в условиях недопустимых входных данных или стрессовых условий окружающей среды, а также способность воспроизводить измерения и результаты.

Надежность создает новые проблемы в контексте систем AI. Архитектуры нейронных сетей представляют собой особую проблему, поскольку их трудно объяснить, а иногда они ведут себя непредсказуемо из-за своей нелинейной природы. Характеристика надежности нейронных сетей является открытой областью исследований, и существуют ограничения как для тестирования, так и для верификации подходов.

Для получения дополнительной информации о надежности нейронных сетей см. [6].

**A.10 Безопасность**

Использование систем AI может представлять новые угрозы безопасности. Безопасность связана с ожиданием того, что система при определенных условиях не приведет к состоянию, в котором жизнь, здоровье, имущество или окружающая среда человека подвергаются опасности. Использование систем AI в автоматизированных транспортных средствах, производственных устройствах и роботах может представлять риски, связанные с безопасностью. Для систем AI в настоящих областях следует учитывать специальные стандарты для определенных областей применения (например, проектирование машин, транспорта, медицинских устройств).

Для получения дополнительной информации о функциональной безопасности в системах AI см. [7]1).

\_\_\_\_\_\_\_\_\_\_\_

1) На этапе подготовки. Этап на момент публикации: ISO/IEC DTR 5469:2022.

**A.11 Безопасность**

Менеджмент риска информационной безопасности определен в [8]. В контексте AI, и, в частности, в отношении систем AI, основанных на подходах МО, следует рассмотреть несколько новых проблем, таких как отравление данных, состязательные атаки и кража моделей, в соответствии с описанием в [3], помимо классических проблем безопасности информации и систем.

**A.12 Прозрачность и объяснимость**

Прозрачность относится как к характеристикам организации, эксплуатирующей системы AI, так и к самим настоящим системам. Организации иногда прозрачны в том, как они применяют такие системы, как они используют собранные данные (например, данные потребителей и пользователей, общедоступные данные, другие собранные наборы данных), какие меры они применяют для управления системами AI, понимают и контролируют свои риски и т. д. Прозрачность систем AI заключается в предоставлении соответствующей информации о системе (например, возможностях и ограничениях) заинтересованным сторонам, чтобы они могли оценить разработку, эксплуатацию и использование систем AI в соответствии со своими целями. Объяснимость системы AI связана со способностью рационализировать и помогать понимать, как для конкретной системы был получен ее результат.

**Приложение B**

*(информационное)*

**Источники риска**

**B.1 Общие сведения**

При определении рисков систем AI следует учитывать различные источники риска в зависимости от характера рассматриваемой системы и контекста ее применения. Источники риска, которые следует учитывать, включают, помимо прочего, проблемы и возможности, описанные в разделах B.2 по B.8.

**B.2 Сложность среды**

Сложность среды [9] системы AI определяет диапазон потенциальных ситуаций, которые система AI должна поддерживать в своем рабочем контексте.

Определенные технологии AI, такие как МО, специально подходят для обработки сложных сред и поэтому часто используются для систем, используемых для сложных сред, таких как автоматизированное вождение. Тем не менее, большой проблемой является определение в процессе проектирования и разработки всех соответствующих ситуаций, которые система должна обрабатывать, и того, чтобы данные обучения и тестирования охватывали все настоящие ситуации.

Следовательно, сложные среды могут привести к дополнительным рискам по сравнению с простыми средами. Особое внимание следует уделить определению степени понимания среды системы AI:

- Полное понимание среды, которое возможно только для простых предсказуемых или контролируемых сред, так что система AI готова ко всем возможным состояниям среды, с которыми она может столкнуться, позволяет лучше контролировать риски.

- В случае частичного понимания из-за высокой сложности или неопределенности среды, так что система AI не может предсказать все возможные состояния среды (например, автономное вождение), нельзя предполагать, что все соответствующие ситуации рассмотрены. Это может привести к уровню неопределенности, который является источником риска и должен учитываться при проектировании таких систем.

**B.3 Отсутствие прозрачности и объяснимости**

Прозрачность заключается в сообщении соответствующим заинтересованным сторонам о соответствующих действиях и решениях организации (например, политик, процессов) и соответствующей информации о системе AI (например, возможностей, производительности, ограничений, вариантов проектирования, алгоритмов, данных обучения и тестирования, процессов и результатов проверки и валидации). Это может позволить заинтересованным сторонам оценить разработку, эксплуатацию и использование систем AI в соответствии со своими ожиданиями. Вид и уровень информации, которая является подходящей, сильно зависят от заинтересованных сторон, варианта использования, типа системы и законодательных требований. Если организации не могут предоставить соответствующую информацию соответствующим заинтересованным сторонам, это может отрицательно повлиять на надежность и подотчетность организации и системы AI.

Объяснимость - это свойство системы AI, заключающееся в том, что важные факторы, влияющие на решение, могут быть выражены таким образом, что люди могут их понять. Модель МО может иметь поведение, которое трудно понять путем проверки модели или алгоритма, используемого для ее обучения, особенно в случае глубокого обучения. Если такие важные факторы не могут быть выражены, это отрицательно влияет на валидацию системы AI и доверие людей к системе, поскольку неясно, почему система приняла решение и может ли она принять правильное решение во всех случаях. Настоящая неопределенность может привести ко многим рискам и сильно повлиять на общие цели, такие как надежность и подотчетность, и на конкретные цели, такие как безопасность, защищенность, справедливость и надежность. Поэтому объяснимость важна не только для заинтересованных сторон как часть прозрачности системы AI, но и для самой организации для ее собственной проверки и верификации системы AI.

Чрезмерная прозрачность и объяснимость также могут привести к рискам, связанным с конфиденциальностью, безопасностью, требованиями конфиденциальности и интеллектуальной собственностью.

**B.4 Уровень автоматизации**

Системы AI могут работать с различными уровнями автоматизации. Они могут варьироваться от отсутствия автоматизации, когда оператор полностью контролирует систему, до полностью автоматизированных систем. Системы AI часто являются автоматизированными системами. В зависимости от конкретного варианта использования автоматизированные решения таких систем могут влиять на различные области, такие как безопасность, справедливость или защита.

Для уровня автоматизации, когда сторонняя организация должна быть готова при необходимости, передача от системы к сторонней организации может быть источником риска (например, временные ограничения, внимание агента).

Для получения дополнительной информации об уровнях автоматизации   
см. ISO/IEC 22989:2022, 5.2.

**B.5 Источники риска, связанные с машинным обучением**

Многие достижения в области AI связаны с МО и его подобластями, такими как глубокое обучение. Режим систем МО критически зависит не только от используемых алгоритмов, но и от данных, на которых обучаются модели МО. Поэтому возможные эффекты на характеристики AI включают:

- Качество данных: качество обучающих и тестовых данных напрямую влияет на функциональность системы. Ненадлежащее качество данных может повлиять на различные цели, такие как справедливость, безопасность и надежность.

- Для систем AI, использующих МО, процессы, используемые для сбора данных, являются источником рисков, которые особенно трудно диагностировать и обнаружить. Например:

- Данные могут стать нерепрезентативными для области применения, что приведет к рискам для бизнес-целей.

- Источники и хранение данных могут повлечь за собой значительные этические и юридические риски. Неспособность защитить процесс сбора данных может привести к рискам от враждебных атак, отравления данных или других манипуляций.

- Системы непрерывного обучения AI направлены на улучшение систем на основе развивающихся производственных данных, в то же время могут усугубить риск, поскольку они могут изменить свое поведение во время использования способом, который не ожидался при вводе в эксплуатацию.

**B.6 Проблемы с аппаратным обеспечением системы**

Источники риска, связанные с аппаратными проблемами, включают, но не ограничиваются следующим:

- Аппаратные ошибки, вызванные неисправными компонентами. Примерами являются короткие замыкания или прерывания одной или нескольких ячеек памяти, неисправные линии шины, дрейфующие генераторы, застрявшие неисправности или паразитные колебания на входах или выходах интегральных схем.

- Программные ошибки, такие как нежелательные временные изменения состояния ячеек памяти или логических компонентов, в основном вызванные высокоэнергетическим излучением.

- Передача обученных моделей МО между различными системами может быть ограничена из-за различных аппаратных возможностей систем с точки зрения вычислительной мощности, памяти и доступности специализированных аппаратных ускорителей AI.

- Когда система AI требует удаленной обработки и хранения, возникают сетевые ошибки, ограничения полосы пропускания и повышенная задержка из-за ограниченного и общего характера сетевых ресурсов.

**B.7 Проблемы жизненного цикла системы**

Неподходящие или недостаточные методы, процессы, а также использование системы AI на протяжении ее жизненного цикла могут привести к рискам. Примерами таких рисков являются:

- Проектирование и разработка: некорректный процесс проектирования может не предвидеть контексты, в которых используется система AI, что приведет к ее неожиданному отказу при использовании в настоящих контекстах.

- Верификация и валидация: неадекватный процесс верификации и валидации для выпуска обновленных версий системы AI может привести к случайным регрессиям или непреднамеренному ухудшению или снижению качества, надежности или безопасности.

- Внедрение: неадекватная конфигурация внедрения может привести к проблемам с ресурсами, связанными с памятью, вычислениями, сетью, хранилищем, избыточностью или балансировкой нагрузки.

- Техническое обслуживание, обновление и пересмотр: система AI, которая больше не поддерживается или не обслуживается разработчиком, но все еще используется, может представлять долгосрочные риски или ответственность для организации-разработчика.

- Повторное использование: функционирующая система AI может использоваться в контексте, для которого она изначально не была разработана, что вызовет проблемы из-за различий в требованиях между спроектированным и фактическим использованием. Например, система, разработанная для идентификации лиц на фотографиях, размещенных в социальной сети, может быть использована для попытки идентификации лиц подозреваемых в совершении преступлений на записях видеонаблюдения, что требует гораздо более высокой степени точности, чем исходный вариант использования.

- Вывод из эксплуатации: организации, которые прекращают использование определенной системы AI или компонента на основе технологий AI, могут потерять информацию или экспертные знания в области принятия решений, предоставленные выведенной из эксплуатации системой. Более того, если для замены выведенной из эксплуатации используется другая система, способ обработки информации и принятия решений организацией может измениться.

**B.8 Технологическая готовность**

Технологическая готовность показывает, насколько развитой является настоящая технология в заданном контексте применения. Менее развитые технологии, используемые при разработке и применении систем AI, могут создавать риски, неизвестные организации или трудно поддающиеся оценке. Для развитых технологий может быть доступно большее разнообразие данных об опыте, что упрощает выявление и оценку рисков. Тем не менее, также существует риск самоуспокоенности и технического долга, если технологии являются развитыми.

**Приложение С**

*(информационное)*

**Менеджмент риска и жизненный цикл системы AI**

Таблица C.1 показывает пример сопоставления между процессами менеджмента риска и жизненным циклом системы AI, в соответствии с определением в   
ISO/IEC 22989:2022.

**Таблица C.1 – Менеджмент риска и жизненный цикл системы AI**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **→**  Менеджмент риска | Структура менеджмента риска AI  (Раздел 5) | Процесс менеджмента риска AI (Раздел 6) | | | | |
| Система AI  Жизненный цикл↓ | Область применения контекст и критерии | Оценка риска | Обработка рисков | Мониторинг и обзор | Учет и отчетность |
| Мероприятия на уровне организации, связанные с управлением рисками | Руководящий орган устанавливает направления для менеджмента риска AI  Высшее руководство принимает обязательства.  Устанавливаются стремление к менеджменту риска высокого уровня и общие критерии | Получаются и обрабатываются отчеты об обратной связи от процессов менеджмента риска систем AI.  В результате, структура управления организационными рисками улучшается за счет расширения и совершенствования инструментов менеджмента риска организации | | | | |
| Каталог критериев риска | Каталог потенциальных источников риска.  Каталог методов оценки и измерения источников риска. | Каталог известных или реализованных мер по смягчению последствий | Каталог известных или реализованных методов мониторинга и управления системами AI. | Каталог установленных методов и определенных форматов для отслеживания, записи, отчетности и обмена информацией о системах AI с внутренними и внешними заинтересованными сторонами. |
| Начало | Руководящий орган изучает цели системы AI в контексте принципов и ценностей организации и заинтересованных сторон.  На основе (обычно многоуровневого) анализа определяет, осуществима ли система AI, и решает проблему, которую организация стремится решить. | Процесс менеджмента риска системы AI и критерии риска системы устанавливаются посредством настройки структуры менеджмента риска  ми организации. | Источники риска, конкретные для системы AI, определяются (возможно, многоуровневым образом) и подробно описываются. | Создается подробный план обработки рисков.  Потенциально определяются методы «доказательства концепции». | Необходимые методы «доказательства концепции» внедряются, тестируются и оцениваются. | Анализ с его результатами и рекомендациями регистрируются и доводятся до сведения высшего руководства. |

*Продолжение таблицы С.1*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **→**  Менеджмент риска | Структура менеджмента риска AI  (Раздел 5) | Процесс менеджмента риска AI (Раздел 6) | | | | |
| Система AI  Жизненный цикл↓ | Область применения контекст и критерии | Оценка риска | Обработка рисков | Мониторинг и обзор | Учет и отчетность |
| Проектирование и разработка | Руководящий орган постоянно переоценивает цели, эффективность и осуществимость системы на основе полученных отчетов обратной связи. | Потенциально критерии риска системы AI изменяются в результате отчетов обратной связи. | Оценка риска выполняется непрерывно (возможно, на нескольких уровнях). | План обработки риска реализуется.  Обработка риска и оценка (остальных) рисков продолжаются до тех пор, пока не будут выполнены установленные критерии риска. | Во время тестирования, верификации и валидации оценивается и корректируется план обработки риска для компонентов системы, а также для всей системы. | Результаты регистрируются и возвращаются в соответствующие мероприятия процесса менеджмента риска.  При необходимости выводы сообщаются цепочке управления и органу управления. |
| Верификация и валидация |
| Эксплуатация, мониторинг | Руководящий орган постоянно переоценивает цели и осуществимость системы на основе полученных отчетов обратной связи. | Потенциально критерии риска системы AI изменяются в результате отчетов обратной связи. | План оценки риска системы потенциально корректируется с учетом изменений критериев риска. | План обработки риска системы потенциально корректируется с учетом изменений рисков в результатах оценки риска. | План обработки риска для компонентов системы оценивается и корректируется |  |
| Постоянная валидация |
| Повторная оценка | Руководящий орган пересматривает цели системы AI и их связь с принципами и ценностями организации и заинтересованных сторон.  На основе анализа определяет, осуществима ли система AI. | Процесс менеджмента риска системы AI и критерии риска системы переоцениваются с учетом любых потенциальных изменений в конкретной цели и области применения системы AI, результатов мониторинга работы и новых нормативных требований. | Список существующих источников риска, специфичных для системы AI, проверяется на предмет релевантности и любых возможных пробелов. | План обработки риска потенциально обновляется.  Обработка риска и (оставшиеся) оценки рисков продолжаются до тех пор, пока не будут выполнены установленные критерии риска. | План обработки риска системы AI переоценивается для внесения необходимых корректировок. |  |
| Вывод из эксплуатации или замена  Запускает новый процесс менеджмента риска с новыми | Руководящий орган пересматривает цели системы AI на основе анализа, определяет, осуществима ли | Устанавливается процесс вывода из эксплуатации менеджмента риска системы AI и критерии риска вывода из | Источники риска, характерные для вывода из эксплуатации системы AI, определяются и подробно | Составляется подробный план обработки рисков. | Необходимые методы «доказательства концепции» внедряются, тестируются и оцениваются. |  |

*Продолжение таблицы С.1*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **→**  Менеджмент риска | Структура менеджмента риска AI  (Раздел 5) | Процесс менеджмента риска AI (Раздел 6) | | | | |
| Система AI  Жизненный цикл↓ | Область применения контекст и критерии | Оценка риска | Обработка рисков | Мониторинг и обзор | Учет и отчетность |
| целями, рисками и их смягчением. | вывод из эксплуатации или замена системы AI. | эксплуатации системы. | описываются. |  |  |  |

**Библиография**

[1] ISO/IEC 38507:2022 Information technology. Governance of IT. Governance implications of the use of artificial intelligence by organizations (Информационные технологии. Стратегическое управление ИТ. Последствия влияния стратегического управления при использовании искусственного интеллекта организациями).

[2] ISO 26000:2010 Guidance on social responsibility (Руководство по социальной ответственности).

[3] ISO/IEC TR 24028:2020 Information technology. Artificial intelligence. Overview of trustworthiness in artificial intelligence (Информационная технология. Искусственный интеллект. Обзор достоверности систем искусственного интеллекта).

[4] ISO/IEC TR 24027:2021 Information technology. Artificial intelligence (AI). Bias in AI systems and AI aided decision making (Информационные технологии. Искусственный интеллект (AI). Необъективность в системах AI и при принятии решений с помощью AI).

[5] ISO/IEC 29134:2017 Information technology. Security techniques. Guidelines for privacy impact assessment (Информационная технология. Методы и средства обеспечения безопасности. Руководящие указания по оценке воздействия на конфиденциальность).

[6] ISO/IEC TR 24029-1:2021 Artificial Intelligence (AI). Assessment of the robustness of neural networks. Part 1. Overview (Искусственный интеллект (AI). Оценка робастности нейронных сетей. Часть 1. Обзор).

[7] ISO/IEC TR 54692) Artificial intelligence. Functional safety and AI systems (Искусственный интеллект. Функциональная безопасность и системы AI).

[8] ISO/IEC 27005:2022 Information security, cybersecurity and privacy protection. Guidance on managing information security risks (Информационная безопасность, кибербезопасность и защита конфиденциальности. Руководство по менеджменту риска нарушения информационной безопасности).

[9] Russell S. J., Norvig P., Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010 (Рассел С. Дж., Норвиг П., Искусственный интеллект: современный подход. 3-е изд. Аппер Сэддл Ривер, Нью-Джерси: Прентис Хол, 2010 г.).

**Приложение В.А**

*(информационное)*

**Сведения о соответствии стандартов ссылочным международным, региональным стандартам, стандартам иностранных государств**

**Таблица В.А.1 – Сведения о соответствии стандартов ссылочным международным, региональным стандартам, стандартам иностранных государств**

|  |  |  |
| --- | --- | --- |
| Обозначение и наименование международного, регионального стандартов, стандарта иностранного государства | Степень соответствия | Обозначение и наименование национального стандарта, межгосударственного стандарта |
| ISO 31000:2018 Risk management. Guidelines (Менеджмент риска. Руководящие указания) | IDT | СТ РК ISO 31000-2020 «Менеджмент риска. Руководящие указания» |
| ISO Guide 73:2009 Risk management. Vocabulary (Менеджмент риска. Словарь) | IDT | СТ РК ИСО 73-2010 «Менеджмент риска. Словарь» |
| ISO/IEC 22989:2022 Information technology. Artificial intelligence. Artificial intelligence concepts and terminology (Искусственный интеллект. Концепции и терминология искусственного интеллекта) | IDT | СТ РК ISO/IEC 22989\* «Информационные технологии. Искусственный интеллект. Концепция и терминология искусственного интеллекта» |
| \*на стадии разработки | | |

|  |
| --- |
| **МКС 35.020 (IDT)** |
| **Ключевые слова:** информационные технологии, искусственный интеллект, руководство, управление рисками, менеджмент риска |

|  |
| --- |
| **МКС 35.020 (IDT)** |
| **Ключевые слова:** информационные технологии, искусственный интеллект, руководство, управление рисками, менеджмент риска |

**РАЗРАБОТЧИК**

РГП на ПХВ «Казахстанский институт стандартизации и метрологии» Комитета технического регулирования и метрологии Министерства торговли и интеграции Республики Казахстан

**Заместитель**

**Генерального директора Е. Амирханова**

**Руководитель**

**Департамента разработки стандартов А. Сопбеков**

**Ведущий специалист**

**Департамента разработки стандартов Ж. Туяков**