ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 72514— 2026 (ИСО/МЭК 42005:2025)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Оценка воздействия системы искусственного интеллекта

(ISO/IEC 42005:2025, Information technology — Artificial intelligence (AI) — AI system impact assessment, MOD)

Издание официальное

Москва
Российский институт стандартизации
2026

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Обществом с ограниченной ответственностью «Институт развития информационного общества» (ООО «ИРИО») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 30 января 2026 г. № 64-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 42005:2025 «Информационные технологии. Искусственный интеллект. Оценка воздействия системы искусственного интеллекта» [ISO/IEC 42005:2025 «Information technology — Artificial intelligence (AI) — AI system impact assessment», MOD] путем внесения технических отклонений, объяснение которых приведено во введении к настоящему стандарту.

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2012 (пункт 3.5).

Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте, приведены в дополнительном приложении ДА

5 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© ISO, 2025
© IEC, 2025
© Оформление. ФГБУ «Институт стандартизации», 2026

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Сокращения 3
5 Разработка и внедрение процесса оценки воздействия системы искусственного интеллекта 3
5.1 Общие положения 3
5.2 Документирование процесса 4
5.3 Интеграция с другими процессами управления организацией 4
5.4 Сроки проведения оценки воздействия системы искусственного интеллекта 4
5.5 Сфера охвата оценки воздействия системы искусственного интеллекта 5
5.6 Распределение ответственности 5
5.7 Установление пороговых значений для высокорискового использования, ограниченного использования и масштабов воздействия 6
5.8 Проведение оценки воздействия системы искусственного интеллекта 7
5.9 Анализ результатов оценки воздействия системы искусственного интеллекта 7
5.10 Документирование и отчетность 7
5.11 Процесс утверждения 8
5.12 Мониторинг и анализ результатов оценки 8
6 Документирование оценки воздействия системы искусственного интеллекта 9
6.1 Общие положения 9
6.2 Область применения оценки воздействия системы искусственного интеллекта 9
6.3 Информация о системе искусственного интеллекта 9
6.4 Информация о данных и их качество 10
6.5 Информация об алгоритме и модели 11
6.6 Среда развертывания 13
6.7 Соответствующие заинтересованные стороны 13
6.8 Фактические и обоснованно предсказуемые воздействия 14
6.9 Меры по устранению вреда и обеспечению пользы 17
Приложение А (справочное) Руководящие указания по использованию настоящего стандарта совместно с *ГОСТ Р ИСО/МЭК 42001* 19
Приложение В (справочное) Руководящие указания по использованию настоящего стандарта совместно с *[2]* 22
Приложение С (справочное) Таксономия вреда и пользы 24
Приложение D (справочное) Совмещение оценки воздействия системы искусственного интеллекта с другими оценками 26
Приложение Е (справочное) Пример шаблона оценки воздействия системы искусственного интеллекта 30
Приложение ДА (справочное) Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте 38
Библиография 39

Введение

Все более широкое применение систем, продуктов, услуг и их компонентов, в которых в той или иной форме используется искусственный интеллект (ИИ), привело к росту обеспокоенности тем, как системы искусственного интеллекта (*СИИ*) могут потенциально воздействовать на все уровни общества. Использование ИИ приносит значительную пользу: автоматизация сложной или опасной работы, более быстрый и точный анализ больших массивов данных, развитие здравоохранения и т. д. Однако существуют опасения относительно обоснованно предсказуемых неблагоприятных последствий использования *СИИ*, включая потенциально вредные, несправедливые или дискриминационные результаты, причинение вреда окружающей среде и нежелательное сокращение рабочих мест.

Разработка и использование *СИИ*, которые, на первый взгляд, не вызывают каких-либо опасений, могут, тем не менее, оказать значительное воздействие (как благоприятное, так и неблагоприятное) на отдельных лиц, группы лиц и общество в целом. Для повышения прозрачности *СИИ* и обеспечения доверия к ним организация, разрабатывающая и использующая эти технологии, может предпринять ряд мер для того, чтобы убедить заинтересованные стороны, подверженные воздействиям *СИИ*, в том, что подобные воздействия были должным образом учтены. Оценки воздействия *СИИ* играют важную роль в рамках более широкой экосистемы усилий в области стратегического управления, оценок рисков и соответствия нормативно-правовым и иным установленным требованиям, которые в совокупности могут создать систему доверия и подотчетности.

Стандарты *[1]*, *[2]* и *ГОСТ Р ИСО/МЭК 42001* формируют важные компоненты этой экосистемы в отношении стратегического управления, оценок рисков и соответствия нормативно-правовым и иным установленным требованиям (через систему менеджмента). В каждом из них подчеркивается необходимость учета воздействия *СИИ* на отдельных лиц и социальные группы. Руководящий орган может учесть это воздействие с целью обеспечения соответствия разработки и применения *СИИ* ценностям и целям организации. Организация, осуществляющая деятельность по управлению рисками, может учесть обоснованно предсказуемые воздействия на отдельных лиц и социальные группы и соответствующим образом включить их в общий процесс оценки рисков организации. Организация, разрабатывающая или использующая *СИИ*, может включить проведение анализа и документирование этого воздействия в свою систему менеджмента, чтобы обеспечить соответствие рассматриваемых *СИИ* ожиданиям соответствующих заинтересованных сторон, а также внутренним и внешним требованиям.

Проведение оценок воздействия *СИИ* и использование ее задокументированных результатов является неотъемлемой частью деятельности по созданию прозрачных и заслуживающих доверия *СИИ* на всех уровнях организации. С этой целью в настоящем стандарте представлены руководящие указания для организаций по внедрению процесса проведения данных оценок и по формированию общего понимания компонентов, необходимых для проведения эффективной оценки.

В настоящий стандарт внесены следующие изменения по отношению к международному стандарту ИСО/МЭК 42005:2025:

- изменены ссылки;

- исключены отдельные положения, которые дублируются по тексту стандарта;

- из библиографии исключена информация о документах, ссылки на которые не использованы при изложении настоящего стандарта;

- включены сокращения, которые выделены в тексте курсивом;

- термины и определения приведены в соответствие с терминологией, установленной в документах по стандартизации Российской Федерации.

Внесение указанных технических отклонений направлено на учет целесообразности использования ссылочных национальных стандартов вместо ссылочных международных стандартов, а также особенностей национальной стандартизации, технологий работы с большими данными и искусственным интеллектом.

Руководящие указания по использованию настоящего стандарта совместно с ГОСТ Р ИСО/МЭК 42001—2024, ИСО/МЭК 23894:2023, таксономия вреда и пользы, совмещение оценки воздействия СИИ с другими оценками приведены в приложениях A—D.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Оценка воздействия системы искусственного интеллекта

Artificial intelligence.
AI system impact assessment

Дата введения — 2026—05—01

1 Область применения

Настоящий стандарт устанавливает руководящие указания для организаций, проводящих оценки воздействия системы искусственного интеллекта на отдельные лица и социальные группы, которые могут быть подвержены воздействию системы искусственного интеллекта и ее предсказуемых применений. Настоящий стандарт устанавливает, как и когда проводить такие оценки и на каких стадиях жизненного цикла системы искусственного интеллекта, а также содержит руководящие указания по документированию оценки воздействия системы искусственного интеллекта.

В настоящем стандарте установлено, как процесс оценки воздействия системы искусственного интеллекта может быть интегрирован в имеющиеся у организации систему управления рисками искусственного интеллекта и систему менеджмента искусственного интеллекта организации.

Настоящий стандарт предназначен для применения организациями, разрабатывающими, предоставляющими или использующими системы искусственного интеллекта. Настоящий стандарт применим в любой организации независимо от масштаба, типа и рода деятельности.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 71476—2024 (ИСО/МЭК 22989:2022) Искусственный интеллект. Концепции и терминология искусственного интеллекта

ГОСТ Р 71484.1 (ИСО/МЭК 5259-1:2024) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 1. Обзор, терминология и примеры

ГОСТ Р 71484.2 (ИСО/МЭК 5259-2:2024) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Показатели качества данных

ГОСТ Р 71484.4 (ИСО/МЭК 5259-4:2024) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 4. Структура процесса управления качеством данных

ГОСТ Р ИСО 31000—2019 Менеджмент риска. Принципы и руководство

ГОСТ Р ИСО 31073—2024 Менеджмент риска. Словарь

ГОСТ Р ИСО/МЭК 42001—2024 Искусственный интеллект. Система менеджмента

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого

Издание официальное

стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены термины по *ГОСТ Р 71476* и *[3]*, а также следующие термины с соответствующими определениями:

3.1

оценка воздействия системы ИИ (AI system impact assessment): Формализованный документированный процесс, посредством которого организация, разрабатывающая, предоставляющая или использующая продукты или услуги с применением ИИ, выявляет, оценивает воздействие на отдельных лиц, группы лиц, социальные группы и принимает соответствующие меры.

[*ГОСТ Р ИСО/МЭК 42001—2024, пункт 3.24*]

3.2 **целевое использование** (искусственный интеллект) (intended use): Использование системы искусственного интеллекта в соответствии с тем, для чего она была спроектирована.

3.3 **нецелевое использование** (искусственный интеллект) (unintended use): Использование системы искусственного интеллекта не по назначению.

3.4

предполагаемые пользователи (intended users): Люди или группа (группы) людей, для которых предназначена продукция.

Примечание — Во многих случаях фактический объем совокупности пользователей отличается от первоначально предполагаемого изготовителем. Характеристика предполагаемой группы пользователей основывается на реалистической оценке того, кто будет фактически пользоваться продукцией.

[*Адаптировано из ГОСТ Р 20282-1—2011, пункт 3.12*]

3.5

заинтересованная сторона (interested party, stakeholder): Лицо или организация, которые могут воздействовать на осуществление деятельности или принятие решения, быть подверженными их воздействию или воспринимать себя в качестве последних.

Примечание — В ИСО/МЭК 22989:2022, 5.19, представлен обзор ролей заинтересованных сторон в области ИИ.

[*ГОСТ Р ИСО/МЭК 42001—2024, пункт 3.2*]

3.6

возможное предсказуемое неправильное использование (reasonably foreseeable misuse): Использование продукта или системы не предназначенным поставщиком образом, но что может быть следствием легко предсказуемого поведения человека.

Примечания

1 Легко предсказуемое поведение человека включает в себя поведение всех типов пользователей, например лиц пожилого возраста, детей и людей с ограниченными возможностями. Для получения большей информации см. [4].

2 В контексте безопасности потребителей термин «возможное предсказуемое использование» все чаще используется в качестве синонима как для термина «предназначенное использование», так и для термина «возможное предсказуемое неправильное использование».

[*Адаптировано из ГОСТ Р 57149—2016, пункт 3.7*]

3.7 **ограниченное использование** (искусственный интеллект) (restricted use): Использование системы искусственного интеллекта, которое ограничено законодательством, политикой организации или договорными обязательствами.

3.8 **высокорисковое использование** (искусственный интеллект) (sensitive use): Использование системы искусственного интеллекта, которое может оказать значительное неблагоприятное воздействие на отдельных лиц, группу лиц или социальные группы.

3.9

высшее руководство (top management): Лицо или группа людей, осуществляющих руководство и управление организацией на высшем уровне.

Примечание 1 — Высшее руководство имеет право делегировать полномочия и предоставлять ресурсы в рамках организации.

Примечание 2 — Если область применения системы менеджмента охватывает только часть организации, под высшим руководством подразумевают тех, кто осуществляет руководство и управляет этой частью организации.

[*ГОСТ Р ИСО/МЭК 42001—2024, пункт 3.3*]

4 Сокращения

В настоящем стандарте применены следующие сокращения:

ИИ — искусственный интеллект;
ИТ — информационные технологии;
МО — машинное обучение;
ПДн — персональные данные;
ПО — программное обеспечение;
СИИ — система искусственного интеллекта;
BIA — оценка воздействия на деятельность;
EIA — оценка воздействия на окружающую среду;
FIA — оценка финансового воздействия;
HRIA — оценка воздействия на права человека;
PIA — оценка воздействия на сохранность ПДн;
SIA — оценка воздействия на безопасность.

5 Разработка и внедрение процесса оценки воздействия системы искусственного интеллекта

5.1 Общие положения

5.1.1 Организация должна применять структурированный и согласованный подход к выполнению и документированию оценки воздействия *СИИ*. Применяемый процесс может варьироваться в зависимости от ряда факторов.

5.1.2 К внутренним факторам относятся:

a) контекст организации, стратегическое управление, цели и задачи, политики и процедуры;
b) договорные обязательства;
c) целевое использование *СИИ*, которая будет разработана или использована;
d) риск-аппетит (*см. ГОСТ Р ИСО 31073—2024, пункт 3.3.27*).

5.1.3 К внешним факторам относятся:

a) действующие нормативно-правовые требования, включая запрещенное использование *СИИ* (*см. [5] и [6]*);
b) политики, руководства и решения регулирующих органов, которые влияют на интерпретацию или соблюдение требований законодательства при разработке и использовании *СИИ*;
c) стимулы и последствия, связанные с целевым использованием *СИИ*;
d) культура, традиции, ценности, нормы и этика в отношении разработки и использования *СИИ*;
e) конкурентная среда и тенденции развития новых продуктов или услуг с применением *СИИ*.

5.2 Документирование процесса

Организация должна документировать весь процесс проведения оценки воздействия *СИИ*. Документацию следует поддерживать в актуальном состоянии и при необходимости предоставлять соответствующим заинтересованным сторонам. В документацию процесса следует включить информацию по аспектам, перечисленным в разделе 5. Предполагаемые результаты документирования процесса могут варьироваться в зависимости от потребностей организации и типа *СИИ* и включать, например, следующее:

a) задокументированные методические руководящие указания;

b) руководство или шаблон для совмещения оценки воздействия *СИИ* с другими оценками (см. приложение D) или отдельный шаблон (см. приложение E);

c) варианты использования для повышения осведомленности и обучения заинтересованных сторон;

d) данные и документы, являющиеся входящими для анализа руководством систем менеджмента, на которые оказывает воздействие ИИ;

e) отчеты о завершенных оценках воздействия *СИИ* и другие материалы, созданные и полученные в процессе оценки.

Документацию следует вести на протяжении всего процесса проведения оценки воздействия *СИИ*, в том числе на стадиях проектирования, перепроектирования, развертывания и оценки. *Сроки хранения документации определяются применимыми нормативно-правовыми и деловыми требованиями, а также политикой организации в отношении сроков хранения документов и данных.*

5.3 Интеграция с другими процессами управления организацией

Организация должна документировать интеграцию процесса оценки воздействия *СИИ* с другими процессами в организации. Документация может охватывать следующие аспекты:

a) как организация интегрирует оценку воздействия *СИИ* с оценкой рисков для организации;

b) как организация интегрирует оценку воздействия *СИИ* с другими видами оценок воздействия;

c) какие процессы стратегического управления организацией, управления рисками и соблюдения нормативных требований существуют или планируются для обработки обоснованно предсказуемых воздействий.

Примечание — Дополнительная информация приведена в приложениях A, B и D.

5.4 Сроки проведения оценки воздействия системы искусственного интеллекта

5.4.1 При разработке процесса оценки воздействия *СИИ* организации следует определить и установить, когда такие оценки следует проводить и на каком уровне, или когда предыдущая оценка воздействия *СИИ* может быть повторно использована, перепрофилирована или пересмотрена, и в каком объеме.

На определение сроков проведения оценок воздействия *СИИ* могут влиять такие факторы, как:

a) применимые требования законодательства;

b) договорные и профессиональные обязательства и обязанности;

c) внутренние структуры, политики, процессы, процедуры и ресурсы, в том числе технологические;

d) уровень риска *СИИ* (дополнительные руководящие указания для организаций приведены в *[2]*, 6.3.4);

e) ожидания соответствующих заинтересованных сторон, в том числе клиентов;

f) внутренние процессы жизненного цикла *СИИ*.

Дополнительные руководящие указания по срокам проведения оценок воздействия *СИИ*, а также по тому, как они могут быть связаны или совмещены с другими оценками воздействия, проводимыми организацией, приведены в приложении D.

5.4.2 Организация должна рассмотреть возможность проведения повторной оценки при изменении таких факторов, как:

a) изменения в целевом использовании *СИИ*, в том числе изменение с точки зрения пользователей *СИИ*;

b) изменение ожиданий клиентов;

c) изменения в самой *СИИ*, в том числе изменение:
1) данных;
2) сложности или типа *СИИ*;
3) производительности *СИИ*;
d) изменения в эксплуатационной среде *СИИ*;
e) изменение в окружении *СИИ*, в том числе:
1) применимых требований законодательства;
2) договорных обязательств;
3) внутренних политик;
4) соответствующих заинтересованных в *СИИ* сторон;
5) местоположений и сфер деятельности, в которых организация ведет или планирует вести свою деятельность.

Оценки рисков ИИ и оценки воздействия *СИИ* следует проводить до внедрения изменений, вызывающих необходимость переоценки.

5.4.3 При определении сроков проведения можно руководствоваться следующим:

a) как часто проводятся оценки воздействия *СИИ*;

b) на какой стадии жизненного цикла *СИИ* выполняется оценка воздействия *СИИ*;

c) как часто обновляется оценка воздействия *СИИ*;

d) при каких обстоятельствах требуется проведение новой оценки воздействия *СИИ* или ее обновление;

e) какие оценки воздействия также связаны с оценкой воздействия *СИИ*.

5.4.4 Организация должна рассмотреть вопрос использования инструментов категоризации при оценке воздействия *СИИ*, например, если организация решает, что оценки воздействия *СИИ* должны проводиться только для *СИИ* «с высоким риском», в рамках процесса оценки следует документировать, что представляет собой *СИИ* «с высоким риском» и что вызывает необходимость проведения оценки воздействия. Процесс категоризации может потребовать проведения оценки воздействия *СИИ* в сокращенном варианте для того, чтобы определить, относится ли данная *СИИ* к категории высокого риска и требуется ли для нее проводить полномасштабную оценку воздействия.

5.5 Сфера охвата оценки воздействия системы искусственного интеллекта

Организация должна определить сферу охвата оценки воздействия *СИИ*, включая сферу применения и границы оценки с учетом внутренних и внешних факторов (см. 5.1), ожиданий соответствующих заинтересованных сторон и обоснованно предсказуемых воздействий на отдельных лиц, группы лиц или социальные группы. Сфера охвата оценки воздействия *СИИ* может включать следующее:

a) всю *СИИ* или ее компонент, который предоставляет функциональные возможности, напрямую доступные пользователям (например, когда в один или несколько компонентов *СИИ* вносятся изменения);

b) описание роли организации в экосистеме ИИ (например, поставщик данных, поставщик моделей, поставщик услуг или продуктов).

Если система состоит из взаимосвязанных *СИИ*, организация должна рассмотреть вопрос о проведении единой оценки воздействия *СИИ*.

Примечание — Изменения в компоненте *СИИ* могут иметь последствия для всей *СИИ* и оказывать влияние на различные заинтересованные стороны.

5.6 Распределение ответственности

Организация должна обеспечить распределение ответственности по оценке воздействия *СИИ* и проинформировать об этом своих сотрудников. Уровень ответственности зависит от нескольких факторов, включая характер оценки воздействия *СИИ*, ее сферу охвата, масштабы, наличие предыдущей оценки, и может включать в себя следующее:

a) установление области применения оценки;

b) обеспечение выделения соответствующих ресурсов;

c) обеспечение доступности документированной информации;

d) установление партнерских отношений с соответствующими заинтересованными сторонами, которые могут внести свой вклад в оценку (например, группы и персонал, ответственные за научно-исследовательскую работу, маркетинг и продажи, безопасность и ведение деятельности, юристы, представители профсоюзов и другие соответствующие заинтересованные стороны);

e) представление отчетности и передача сведений о результатах оценки, а также иной соответствующей информации другим процессам, службам и/или структурным подразделениям организации;

f) установление процедур согласования/утверждения и передачи рассмотрения на более высокий уровень;

g) обеспечение доступности оценки воздействия *СИИ* для соответствующего анализа со стороны руководства в рамках системы менеджмента ИИ и организации.

При распределении обязанностей следует учитывать наличие необходимого опыта и компетенций, значимость конкретных ролей и виды информации, доступ к которой необходим исполнителям этих ролей.

При распределении ответственности организация должна учитывать необходимость междисциплинарного подхода для рассмотрения воздействий, которые возможно предвидеть и реагировать на них на протяжении всего жизненного цикла *СИИ*.

Примечание — К релевантным дисциплинам относятся инженерия, охрана труда и техника безопасности, права человека, этика и социальные науки, такие как социология и психология.

5.7 Установление пороговых значений для высокорискового использования, ограниченного использования и масштабов воздействия

Важнейшей частью процесса оценки воздействия *СИИ* является документирование пороговых значений, особенно при использовании *СИИ*. Организация должна установить пороговые значения, исходя из контекста своей деятельности. При этом следует принимать во внимание следующие аспекты:

a) применимые требования законодательства;

b) ожидания соответствующих заинтересованных сторон;

c) текущее состояние;

d) польза от применения *СИИ*;

e) культурные, трудовые и общественные нормы;

f) применимые этические подходы к использованию ИИ.

В зависимости от типов и количества пороговых значений организация может внедрить дополнительные процессы, включая процессы анализа и согласования. Например, организация может установить, что определенное использование или обоснованно предсказуемое нецелевое использование *СИИ* являются высокорисковыми или запрещены политикой организации. Требования законодательства и иные внешние требования могут дополнительно устанавливать высокие риски использования или обоснованно предсказуемого нецелевого использования *СИИ*. Если оценка воздействия *СИИ* показывает, что ее планируемое использование категорируется как «высокорисковое» или «ограниченное», организация должна задокументировать, какими в таком случае будут последующие шаги, выполняемые в рамках общего процесса.

Примеры

1 Примером случая высокорискового использования могут служить СИИ, предназначенные для автоматизации принятия решений о кредитовании, поскольку такие системы могут оказать значительные неблагоприятные финансовые воздействия на отдельных лиц.

2 Если установлено, что определенные способы использования являются высокорисковыми или должны ограничиваться, то вопрос об их использовании может быть передан на рассмотрение руководству более высокого уровня (в том числе для согласования/утверждения). Рассматривают СИИ, которая оказывает значительное воздействие на права человека (в особенности на права детей), неблагоприятные последствия которого невозможно смягчить с помощью одних только технических усовершенствований. Такие вопросы могут быть переданы для рассмотрения на более высоком уровне, чтобы решить, стоит ли перепроектировать данную СИИ из-за ее неблагоприятного воздействия.

Организация также должна учитывать, как определяются общие масштабы воздействия и измеряются обоснованно предсказуемые воздействия. Результаты оценки воздействия *СИИ* могут указывать, например, на то, что целевое использование *СИИ* не является высокорисковым, однако может затронуть большую группу пользователей.

Организация должна учитывать этот факт при определении масштаба воздействия и выборе дальнейших действий в процессе оценки воздействия *СИИ*. Организации следует учитывать масштабы обоснованно предсказуемых воздействий при определении различных компонентов оценки воздействия *СИИ*. Организация должна разработать критерии для типов обоснованно предсказуемых воздействий, которые могут быть результатом работы *СИИ*. Например, для оценки степени обоснованно предсказу-

емых воздействий оценка воздействия *СИИ* может включать критерии ранжирования оценки воздействий как очень серьезных, если они могут привести к значительным неблагоприятным последствиям для организации, групп лиц или социальных групп.

5.8 Проведение оценки воздействия системы искусственного интеллекта

При проведении оценки воздействия *СИИ* организация должна учитывать (в зависимости от ситуации):

a) выявление обоснованно предсказуемых воздействий *СИИ* (например, воздействие на права человека и основные права и свободы) для всех затронутых лиц, группы лиц и социальных групп. Следует учитывать как благоприятные, так и неблагоприятные воздействия;

b) дополнительные аспекты (не ограничиваясь ими):

1) трансформирование таких воздействий в риски для организации;

2) воздействие многочисленных целей и задач организации, связанных с ИИ, на группы лиц и социальные группы;

c) периодическое или ситуативное выявление воздействия на отдельных лиц и группы лиц, например, как можно раньше и на определенных стадиях жизненного цикла *СИИ*;

d) оценку величины и вероятность выявленных воздействий и определение, как будет осуществляться реагирование на данное воздействие в рамках процесса управления рисками.

Рекомендуется поощрять и интегрировать в оценку вклад представителей различных сообществ, поскольку это может обеспечить разнообразие точек зрения.

5.9 Анализ результатов оценки воздействия системы искусственного интеллекта

Результаты оценки воздействия *СИИ* могут играть ключевую роль в ответственном использовании и развитии *СИИ*. Данные результаты следует анализировать и учитывать при принятии как технических, так и управленческих решений.

На техническом уровне анализ результатов оценки воздействия *СИИ* допускается использовать для повышения качества продукта или услуги, встраивания механизмов защиты от нецелевого использования или обоснованно предсказуемого злонамеренного использования, повышения безопасности и робастности *СИИ* и т. д. Анализ результатов также может включать в себя определение типов мер для обеспечения пользы и устранения вреда, как описано в 6.8.2. Кроме того, результаты оценки следует сравнить с пороговыми значениями, установленными в 5.7, и, если пороговые значения не соблюдены, следует разработать план действий по исправлению ситуации.

На управленческом уровне анализ может использоваться как исходная информация о предвидимых воздействиях *СИИ* в процессе управления рисками организации. Данный процесс может включать рассмотрение оценок воздействия отдельных *СИИ*, комплексный анализ нескольких оценок для выявления тенденций, способных повлиять на риски, цели и задачи организации в целом.

5.10 Документирование и отчетность

5.10.1 Организация должна определить и задокументировать требования к ведению учета оценки воздействия *СИИ*, а также требования к документированию результатов оценки воздействия *СИИ* и мер реагирования на выявленные воздействия. Порядок документирования процесса оценки может зависеть от потребностей организации. Он может быть реализован посредством использования соответствующих шаблонов документов, функциональных возможностей системы или другим способом, обеспечивающим надлежащее документирование и подготовку отчетности.

5.10.2 У организации могут иметься как внутренние потребности в отчетности, так и потребности в отчетности перед внешними сторонами. Внутренняя отчетность может включать в себя отчетность перед:

a) высшим руководством;

b) сотрудниками и привлеченными специалистами или, при необходимости, их представителями;

c) производителями ИИ;

d) партнерами ИИ.

5.10.3 Отчетность перед внешними заинтересованными сторонами (внешняя отчетность) должна соответствовать тому, кем являются эти стороны, их ожиданиям и потребностям в прозрачности, — исключая при этом раскрытие информации, являющееся ненадлежащим с точки зрения обеспечения без-

опасности, конфиденциальности и защиты персональных данных. Внешняя отчетность может включать в себя отчетность перед:

a) соответствующими уполномоченными органами, включая директивные и регулирующие органы;

b) клиентами ИИ;

c) партнерами ИИ;

d) отдельными лицами, группами лиц или социальными группами, подвергающимися воздействию *СИИ*.

5.10.4 Отчетность об оценке воздействия *СИИ* перед соответствующими заинтересованными сторонами может включать:

a) информацию о фактическом или обоснованно предсказуемом воздействии, связанную с целевыми вариантами использования *СИИ*, а также о другом потенциально полезном использовании и обоснованно предсказуемом нецелевом использовании (по крайней мере, в общем плане);

b) адекватное количество аспектов выявленных воздействий, в том числе максимально возможное (в разумных пределах) количество типов воздействий, описанных в 6.8;

c) меры, принятые для реагирования на выявленные воздействия.

Сообщая информацию об обоснованно предсказуемом нецелевом использовании, организация может опустить детали, которые могут быть использованы злоумышленниками.

5.10.5 Отчеты следует структурировать и формулировать таким образом, чтобы в них предоставлялась информация, необходимая соответствующим заинтересованным сторонам для оценки, при необходимости, соответствия *СИИ* требованиям законодательства, связанным с оценкой воздействия. Отчеты следует структурировать и формулировать таким образом, чтобы в тех случаях, когда некоторые части недоступны читателю для защиты интеллектуальной собственности или обеспечения безопасности, они оставались понятными.

5.11 Процесс утверждения

Организация должна документировать все решения о согласовании, разрешении и утверждении, которые потребуется принять в рамках процесса оценки воздействия *СИИ*. К данным решениям относятся разрешение превышений установленных пороговых значений (см. 5.7), а также окончательное утверждение итоговых результатов оценки воздействия *СИИ*. Могут быть приняты во внимание следующие аспекты:

a) ситуации, в которых возникает потребность в управленческом решении (например, при превышении порогового значения или по завершении оценки воздействия *СИИ*);

b) лицо, ответственное за принятие управленческого решения (может зависеть от типа решения);

c) потребность во внешнем утверждении (например, регулирующими органами, внешними соответствующими заинтересованными сторонами).

5.12 Мониторинг и анализ результатов оценки

5.12.1 Организация должна определить и задокументировать процедуры мониторинга и пересмотра оценки воздействия *СИИ*, включая:

a) периодичность проведения (с установленной периодичностью и/или в качестве реакции на какие-либо события-триггеры);

b) ответственных за мониторинг и пересмотр;

c) состав мероприятий, входящих в мониторинг и пересмотр.

5.12.2 В результаты анализа оценки *СИИ* следует включать решения, касающиеся:

a) возможностей для непрерывного совершенствования и потребностей в изменениях в планировании и проведении оценки воздействия *СИИ*;

b) изменений в процессах оценки и обработки рисков;

c) изменений в целях и задачах использования ИИ;

d) изменений пороговых значений для ранжирования по степени воздействия при высокорисковом или ограниченном использовании.

6 Документирование оценки воздействия системы искусственного интеллекта

6.1 Общие положения

В настоящем разделе содержатся руководящие указания, которые могут быть включены в документацию по оценке воздействия *СИИ*. Организация должна определить потребности, исходя из контекста своей деятельности. Не все руководящие указания, приведенные в данном разделе, применимы к любой организации. Например, организация, внедряющая *СИИ*, разработанную третьей стороной, не располагает подробной информацией о данных (6.4) или алгоритме (6.5).

6.2 Область применения оценки воздействия системы искусственного интеллекта

В области применения оценки воздействия *СИИ* следует описать рассматриваемую *СИИ*, ее предсказуемые воздействия, а также внутренние и внешние факторы, влияющие на эти воздействия. В описание *СИИ* при оценке ее воздействия следует включить:

a) идентификацию рассматриваемой *СИИ*;

b) идентификацию отдельных лиц, групп лиц и социальных групп, которые могут подвергнуться воздействию *СИИ*, включая их потребности и ожидания, релевантные для оценки воздействия *СИИ* обоснованно идентифицируемые организацией;

c) описание целевого использования и категорий вариантов использования;

d) перечень способов обоснованно предсказуемого нецелевого использования, которые могут повлиять на идентификацию отдельных лиц, групп лиц и социальных групп.

При необходимости в описании области применения следует четко указать исключения из перечисленных пунктов a)—d).

Область применения оценки воздействия может включать или дополняться другими оценками воздействия (например, оценкой воздействия на права человека или процедурами должного соблюдения прав человека, см. приложение D), основанными на различных частях информации о *СИИ* (как описано в 6.3).

6.3 Информация о системе искусственного интеллекта

6.3.1 Описание системы искусственного интеллекта

В оценку воздействия *СИИ* следует включить базовое описание *СИИ*, в котором говорится, что делает *СИИ*, и приводится высокоуровневое описание принципа ее работы. В описании следует указать, какими возможностями обладает *СИИ*, чтобы предоставить потенциальным рецензентам необходимый контекст для понимания *СИИ* и окружения, в котором она функционирует. Описание *СИИ* может содержать дополнительные сведения, определяемые организацией, например:

a) обзор архитектуры *СИИ*;

b) технические требования и спецификации;

c) демонстрационные материалы и доказательства правильности концепций.

6.3.2 Функции и возможности системы искусственного интеллекта

В документации по функциям *СИИ* следует изложить более конкретное и подробное описание функций и возможностей, а также информацию о том, имеются ли они сейчас или являются запланированными. В документацию может входить, например, высокоуровневое описание:

a) прогнозов, типов данных, алгоритмов и моделей, используемых *СИИ* (см. 6.4 и 6.5);

b) взаимодействия с пользователем;

c) конфигурации аппаратного обеспечения и *ПО*;

d) предупреждений и пороговых значений.

При документировании функций и возможностей *СИИ* организации следует идентифицировать технические подробности и ограничить их раскрытие определенным группам лиц, чтобы снизить вероятность нецелевого использования.

Другая документация, которую следует рассмотреть в связи с функциями и возможностями *СИИ*, — это зависимость от других систем и их функций, и возможностей; например, если документируемая *СИИ* опирается на модель, которая является частью другого продукта или системы.

6.3.3 Назначение системы искусственного интеллекта

Назначение *СИИ* следует задокументировать, чтобы помочь потенциальным рецензентам понять, почему организация создает или использует *СИИ* и как технология ИИ способствует достижению этих целей и задач. При документировании назначения *СИИ* организация должна учитывать:

a) конечного пользователя или основного клиента *СИИ*;

b) как *СИИ* удовлетворяет потребности пользователей;

c) ценностное предложение *СИИ*;

d) компромиссы, на которые пришлось пойти с связи с принятием решения об использовании *СИИ*.

6.3.4 Целевое использование

Влияние *СИИ* на отдельных людей, группы людей и социальные группы частично зависит от ее целевого использования. Целевое использование можно конкретизировать в виде вариантов использования или сценариев, которые могут включать:

a) *цели и задачи СИИ* для определенных групп конечных пользователей или клиентов;

b) *условия использования СИИ конечными пользователями (место и время);*

c) сведения о том, имеются ли ограничения для *СИИ* (включая временные ограничения, например, период времени, в течение которого *СИИ* должна работать).

У *СИИ*, а также у функций и возможностей *СИИ* может быть разнообразное целевое использование. Организация должна выявить и задокументировать варианты целевого использования, отразив, в том числе, полезные с ее точки зрения.

Пример — СИИ, разработанная банком, которая может идентифицировать человека по биометрическому сканированию его лица. Целевое использование может включать:

- предоставление доступа к мобильному приложению банка;

- обеспечение индивидуального подхода в отделении банка.

6.3.5 Нецелевое использование

6.3.5.1 Организация должна идентифицировать и документировать обоснованно предсказуемое нецелевое использование.

Пример — СИИ, разработанная банком, которая может идентифицировать человека по биометрическому сканированию его лица. Нецелевое использование может включать неправомерное использование ПДн.

Примечания

1 Некоторые виды нецелевого использования можно найти в справочных источниках о методах эксплуатации в сфере ИИ, о типах сбоев и типах угроз в системах МО, об идентификации и моделировании угроз в *СИИ*.

2 В *ГОСТ Р ИСО/МЭК 42001* приведены руководящие указания по информированию соответствующих заинтересованных сторон о приемлемом, неприемлемом и запрещенном использовании.

6.3.5.2 Для определения дополнительного потенциально полезного использования *СИИ* организация должна применять методы, позволяющие:

a) выявлять другое полезное использование *СИИ*, которое может реализовать организация;

b) выявлять другое полезное использование *СИИ*, которое может реализовать пользователь.

6.4 Информация о данных и их качество

6.4.1 Общие положения

Некоторые *СИИ* используют обучающие, проверочные и тестовые наборы данных для создания, тестирования и контроля обученных моделей. Эксплуатационные данные применяются к обученным моделям для создания выходных данных. Использование обучающих, проверочных и тестовых наборов данных, не соответствующих требованиям в контексте *СИИ*, может привести к получению неверных результатов, которые впоследствии окажут неблагоприятные воздействия на отдельных лиц, группы лиц и социальные группы. Кроме того, прошлая или текущая дискриминация и несправедливость в отношении групп лиц, отраженная в обучающих наборах данных, может привести к тому, что обученная модель будет распространять или даже усиливать дискриминацию или несправедливость.

Примечание — Более подробная информация о типах данных, используемых в *СИИ*, приведена в *ГОСТ Р 71476*.

6.4.2 Информация о данных

Наборы данных могут сопровождаться описательной информацией разной степени полноты — от ее отсутствия до исчерпывающей. Организация может определить воздействие использования конкретного набора данных в *СИИ*, когда информация является исчерпывающей. Организация должна это учитывать при определении того, какая информация должна быть включена в оценку воздействия *СИИ*.

При документировании наборов данных, используемых в *СИИ*, в рамках оценки воздействия *СИИ* организация должна учитывать процессы обеспечения качества используемых данных, стратегического управления этими данными, обеспечения безопасности и защиты ПДн. Сюда следует включить следующее:

a) доступность информации о данных для потенциального анализа человеком. Информация о данных может храниться в машиночитаемой форме, но также должна быть доступна для анализа рецензентом;

b) способ и источник получения данных (соответствие процесса комплектования данных правовым обязательствам, этическим принципам и другим политикам организации);

c) количество данных по отношению к потребностям *СИИ*;

d) происхождение набора данных (например, сведения о создателях данных, распорядителях данных, пользователях данных, процессы, применяемые к данным, или местоположения хранилищ данных);

e) известные или обоснованно предсказуемые риски нежелательной предвзятости в наборе данных;

f) географические регионы, охватываемые набором данных;

g) процессы обеспечения качества данных;

h) контроль доступа к набору данных и его защита;

i) временны́е рамки данных (т. е. начальная и конечная дата добавления данных в набор);

j) анализ данных соответствующими заинтересованными сторонами (в том числе, были ли проведены разноплановые анализы);

k) являются ли наборы данных реальными, синтетическими или полусинтетическими.

Примечание — Дополнительная информация, которая может сопровождать набор данных, приведена в *[7]*. Информация о защите ПДн приведена в 6.8.2.5. Более подробная информация о нежелательной предвзятости приведена в *[8]*.

6.4.3 Документирование качества данных

При документировании качества данных для оценки воздействия *СИИ* организация должна учитывать следующее:

a) какие характеристики качества данных применимы в данном контексте;

b) требования к качеству данных, которые соблюдены;

c) планы по достижению требований к качеству данных, которые еще не соблюдены;

d) использование жизненного цикла качества данных при разработке *СИИ*;

e) утвержденное соответствующими заинтересованными сторонами использование набора данных в *СИИ*.

Примечание — Информация о терминах, концепциях и жизненных циклах качества данных приведена в *ГОСТ Р 71484.1*. Информация о модели, характеристиках и мерах обеспечения качества данных приведена в *ГОСТ Р 71484.2*. Информацию о процессах обеспечения качества данных см. в *ГОСТ Р 71484.4*.

6.5 Информация об алгоритме и модели

6.5.1 Общие положения

Организация должна документировать информацию об алгоритмах и моделях (включая их разработку), чтобы обеспечить возможность рассмотрения и оценки связанных с ними обоснованно предсказуемых воздействий.

Примечание — Более подробная информация об алгоритмах и моделях МО приведена в *ГОСТ Р 71476* и *[3]*.

В некоторых случаях, например для лица, производящего развертывание, или пользователя, некоторая информация, рекомендованная в 6.5.2—6.5.5, может быть недоступна, доступна в разной степени детализации или может не считаться необходимой. Однако такая информация необходима для

понимания компонентов *СИИ*. Информация о различных компонентах, таких как алгоритмы и модели, может быть полезной. Необходимо, чтобы оценка воздействия *СИИ* проводилась в отношении всей *СИИ* в целом.

6.5.2 Информация об алгоритмах, используемых организацией

При разработке документации для алгоритмов, используемых в *СИИ*, организация должна учитывать следующее:

a) уместность выбранных алгоритмов по отношению к целям деятельности и самой задаче ИИ;

b) происхождение алгоритмов и модификаций (например, научные статьи, журналы и презентации или репозитории исходного кода);

c) валидность алгоритма (например, был ли алгоритм протестирован и признан авторитетными экспертами в данной области);

d) предыдущие виды использования алгоритмов на практике и их эффективность (например, были ли *СИИ*, основанные на алгоритмах, введены в эксплуатацию или остановлены, был ли ранее причинен вред в результате использования алгоритмов, каковы показатели эффективности в предыдущих случаях использования);

e) требования к данным алгоритмов;

f) подверженность алгоритмов нежелательным выводам (например, ложным корреляциям, переобучению или распространению нежелательной предвзятости).

6.5.3 Информация о разработке алгоритма

При разработке документации по алгоритмам в рамках оценки воздействия *СИИ* организация должна учитывать следующее:

a) соблюдение применимых нормативно-правовых, организационных или технических требований;

b) планы по достижению требований, которые еще не соблюдены;

c) документирование алгоритмов на протяжении всего жизненного цикла разработки *СИИ*;

d) процесс утверждения алгоритма, используемого в *СИИ*.

6.5.4 Информация о моделях, используемых в системе искусственного интеллекта

При разработке документации для моделей, используемых в *СИИ*, организация должна учитывать следующее:

a) данные (например, обучающие, проверочные, тестовые) и алгоритмы, использованные для разработки моделей (см. 6.5.3);

b) методы обучения модели;

c) установление факта отсутствия повторов данных среди обучающих, тестовых и проверочных наборов данных;

d) шаги, предпринятые для обучения, тестирования и валидации нескольких моделей;

e) критерии и шаги, предпринятые для выбора подходящих признаков;

f) критерии выбора модели;

g) показатели, используемые для оценки производительности моделей (например, точность, рабочая характеристика приемника, матрица ошибок, среднеквадратическая ошибка или среднее расстояние до центра кластера);

h) оценку того, насколько хорошо выбранные модели обобщают эксплуатационные данные;

i) оценку того, сохраняют или создают ли выбранные модели нежелательную предвзятость или другой вред;

j) оценку того, содержат ли выходные данные модели или создают ПДн;

k) робастность модели;

l) обнаружение и исправление дрейфа данных (например, переобученность модели, если эксплуатационные данные начинают отличаться по характеристикам от исходных обучающих данных);

m) критерии и процесс переобучения модели;

n) воздействия непрерывного обучения (если применимо);

o) вычислительные и экологические соображения.

Для моделей, дообученных (например, доработанных) организацией, этот перечень рассматривают с учетом новых данных и сценариев. При рассмотрении моделей перед дообучением рекомендуется ознакомиться с информацией, предоставленной поставщиками моделей.

Для предварительно обученных моделей, импортированных и используемых в *СИИ*, перечисления e)—i), а также тестирование в перечислениях a), c) и d) следует рассматривать с учетом новых данных и сценариев.

6.5.5 Информация о разработке модели

При разработке документации по моделям в рамках оценки воздействия *СИИ* организация должна учитывать следующее:

a) соблюдение применимых нормативно-правовых, организационных или технических требований;

b) планы по достижению требований, которые еще не соблюдены;

c) документирование моделей на протяжении всего жизненного цикла разработки *СИИ*.

6.6 Среда развертывания

6.6.1 Географический регион и языки

Воздействия *СИИ* могут быть связаны с географическим, культурным или законодательным контекстом стран или регионов, где они развернуты, поэтому их следует учитывать при документировании оценок воздействия *СИИ*.

При документировании воздействия *СИИ*, развернутой в конкретной стране или регионе, организация должна учитывать:

a) где *СИИ* развернута в настоящее время и где она может быть развернута в будущем;

b) особые требования законодательства в географическом регионе развертывания;

c) культурные особенности географического региона развертывания;

d) дискриминируемые группы, группы риска и маргинализированные группы;

e) языки, на которых говорят в географическом регионе развертывания, а также, в частности, в системах обработки естественного языка, какие языки поддерживаются в настоящее время;

f) поведенческие, социальные и физиологические особенности человека, которые могут быть представлены в географическом регионе.

Организация также должна документировать планируемые действия по совершенствованию будущих версий *СИИ* с учетом географических и культурных особенностей или требований законодательства страны или региона развертывания.

Для обеспечения надлежащего представительства географически сконцентрированных групп с общими языками, культурами или особенностями *СИИ* следует обучить, протестировать и проконтролировать на основе данных, характерных для местности, где она будет развернута. Если *СИИ* предварительно обучена, ее следует тщательно протестировать и проконтролировать, используя данные, характерные для местности, где она будет развернута.

6.6.2 Сложность и ограничения среды развертывания

Наряду с географической средой развертывания, в документацию оценки воздействия *СИИ* следует включить информацию о технической среде развертывания и соответствующих ограничениях. В документацию может входить такая информация о среде, как:

a) сведения о том, является ли она онлайн-сервисом, приложением и т. д.;

b) сведения о том, является ли она платформой;

c) сведения о том, есть ли облачное развертывание;

d) особенности, связанные с открытым исходным кодом;

e) сведения о том, доступна ли она только на определенном оборудовании;

f) сведения о том, доступна ли она только в составе определенных приложений;

g) есть ли инфраструктура, необходимая для *СИИ*;

h) требования обеспечения безопасности;

i) соответствующий контекст, включая знания о *СИИ* и предыдущий опыт работы с ней.

6.7 Соответствующие заинтересованные стороны

6.7.1 Общие положения

В процессе документирования оценки воздействия организация должна учитывать конкретное воздействие на отдельных лиц, группы лиц и социальные группы, а также роль и обязательства перед другими заинтересованными сторонами.

6.7.2 Непосредственно подверженные влиянию заинтересованные стороны

При оценке воздействия *СИИ* организация должна выявлять и документировать лица, группы лиц или социальные группы, которые могут подвергнуться влиянию *СИИ* и ее целевому использованию. К соответствующим заинтересованным сторонам относятся как те, кто непосредственно взаимодейству-

ет с *СИИ*, так и те, кто может подвергаться ее влиянию без необходимости взаимодействовать с ней. Некоторые примеры соответствующих заинтересованных сторон:

a) пользователи ИИ;
b) уязвимые лица;
c) рабочие;
d) субъекты данных.

Группы лиц и социальные группы, которые могут подвергнуться воздействию *СИИ*. В процессе оценки воздействия *СИИ* следует консультироваться с отдельными лицами или группами лиц, представляющими соответствующие заинтересованные стороны. Результаты консультаций следует документировать.

Масштаб и предмет консультаций с соответствующими заинтересованными сторонами должны быть пропорциональны результату и оценке первоначального уровня риска, полученным в результате процесса категоризации, соизмеримого с масштабом и предметом процесса оценки воздействия *СИИ* (например, если оценка показывает высокий уровень риска, то оценка воздействия *СИИ*, а также процесс консультаций могут быть более робастными). Вне зависимости от уровня риска, процесс консультаций может позволить организации понять интересы и опасения соответствующих сторон и добросовестно на них отреагировать. Консультации должны предоставлять заинтересованным сторонам своевременную и релевантную информацию, включать различные точки зрения, должным образом учитывая особые потребности уязвимых групп или отдельных лиц, добиваясь, чтобы испытывающие воздействие соответствующие заинтересованные стороны не подвергались преследованию и противодействию, сохраняя конфиденциальность и анонимность, где это необходимо, и предоставляя отдельным лицам возможность оспаривать решения и, в случае необходимости, обращаться за компенсацией.

6.7.3 Другие заинтересованные стороны

При оценке воздействия *СИИ* организация должна учесть вопрос об идентификации и документировании других заинтересованных сторон. данные заинтересованные стороны необязательно подвергаются воздействию *СИИ*, но имеют к ней отношение, которое необходимо зафиксировать. Некоторые примеры включают:

a) производителей ИИ;
b) соответствующие уполномоченные органы, такие как директивные и регулирующие органы;
c) профессиональные ассоциации и консорциумы.

6.8 Фактические и обоснованно предсказуемые воздействия

6.8.1 Общие положения

Для заинтересованных сторон, указанных в 6.7.2, следует идентифицировать обоснованно предсказуемые пользу и вред.

6.8.2 Польза и вред

6.8.2.1 Общие положения

Для каждой заинтересованной стороны следует проанализировать обоснованно предсказуемые пользу и вред. Например, организация, проводящая оценку воздействия, должна учитывать обоснованно предсказуемую пользу, которую каждая заинтересованная сторона может ожидать в результате взаимодействия с *СИИ*. Подразделы данного раздела допускается использовать для учета обоснованно предсказуемых пользы и вреда, связанных с различными целями и задачами. Эти цели и задачи не являются исчерпывающими, другие также могут быть рассмотрены. Более подробная информация приведена в приложении С.

6.8.2.2 Подотчетность

Подотчетность — это состояние или обязанность физического или юридического лица отвечать за действия, решения или результаты работы, включая решения, принимаемые *СИИ.*

Использование *СИИ* может изменить существующие системы подотчетности. Дополнительная информация о цели подотчетности приведена в *[2]* (пункт А.2), о подотчетности при использовании ИИ организациями с точки зрения управления организацией приведена в *[1]*.

Для каждой идентифицированной заинтересованной стороны организация должна проанализировать обоснованно предсказуемые пользу и вред, связанные с подотчетностью в результате использования *СИИ, например:*

a) воздействия на заинтересованную сторону, включая следующие аспекты:

1) возможность того, что использование *СИИ* может повлиять на их правовой статус, возможности в жизни или права человека;

2) возможность нанесения *СИИ* физического или психологического вреда заинтересованной стороне;

b) воздействия на заинтересованную сторону в случае, если *СИИ* не выполняет свое назначение должным образом;

c) воздействия на заинтересованную сторону в результате отбора и использования данных, применяемых для разработки *СИИ*;

d) воздействия на заинтересованную сторону в результате отсутствия надлежащего человеческого надзора за *СИИ* и ее результатами;

e) воздействия на заинтересованную сторону в результате невозможности отменить или оспорить решения, принятые *СИИ*.

6.8.2.3 Прозрачность

Прозрачность — это свойство организации или системы сообщать заинтересованным сторонам надлежащую информацию о действиях, решениях и свойствах системы в исчерпывающем, понятном и доступном виде. Для *СИИ* информация о прозрачности системы может включать такие аспекты, как характеристики, показатели деятельности, ограничения, компоненты, процедуры, показатели, цели проектирования, проектные решения и допущения, источники данных и протоколы разметки.

Объем информации и понимание *СИИ* потенциально могут оказать воздействие на заинтересованные стороны. Дополнительная информация о цели подотчетности приведена в *[2]* (пункт А.12).

Для каждой идентифицированной заинтересованной стороны организация должна проанализировать обоснованно предсказуемые пользу и вред, связанные с прозрачностью при использовании *СИИ*.

Примеры воздействия на заинтересованную сторону, которое следует учесть:

a) достаточно ли информации о возможностях *СИИ*, ее ограничениях и сдерживающих факторах, предоставлена ли информация в понятном для заинтересованных сторон формате;

b) осознают ли они, что взаимодействуют с *СИИ*.

6.8.2.4 Справедливость и дискриминация

Значение понятия «справедливость» в большой степени зависит от контекста, а также от ряда других факторов, включая цели организации, назначение, для которого используется *СИИ*, и окружение, в котором она работает. Использование *СИИ* для автоматизированного принятия решений может быть несправедливым по отношению к конкретным лицам или группам лиц. Несправедливые результаты могут быть получены в результате ряда причин, таких как предвзятость объективных функций, несбалансированные наборы данных, отражение человеческих предубеждений при подготовке обучающих данных и предоставлении системам обратной связи. Несправедливость также может быть вызвана предвзятостью в концепции продукта, постановке задачи или выборе времени и места развертывания *СИИ*. Дополнительную информацию о задаче обеспечения справедливости см. в *[2]*, приложение А.6.

Для каждой идентифицированной заинтересованной стороны организации следует проанализировать обоснованно предсказуемые пользу и вред, связанные с задачей организации по обеспечению справедливости, которые используемые *СИИ* могут принести отдельным лицам и социальным группам.

Ниже приведены примеры воздействия на заинтересованную сторону, которое следует учесть:

a) производительность *СИИ* становится неоправданно лучше или хуже для конкретной группы населения (например, в зависимости от возраста, пола или религиозной принадлежности);

b) *СИИ* создает такой образ заинтересованной стороны, который несправедливо дискриминирует ее по признаку ассоциированных с ней групп населения;

c) аспекты доступности для заинтересованных лиц, чтобы они не были вытеснены из-за инвалидности или отсутствия цифровых навыков;

d) влияние нежелательной предвзятости на группы риска. Дополнительная информация приведена в *[9]*.

6.8.2.5 Защита персональных данных

Если *СИИ* при обработке, во входных или выходных данных оперируют ПДн, то есть вероятность, что система ИИ может оказать воздействие на их защиту. В *[2]* (пункт А.8) содержится информация о задачах защиты ПДн, включая их сбор, использование, раскрытие и вывод, связанных с *СИИ*.

Для каждой идентифицированной заинтересованной стороны организация должна проанализировать обоснованно предсказуемые пользу и вред, связанные с защитой ПДн при использовании *СИИ*. Субъекты ПДн могут являться заинтересованной стороной.

Примеры воздействия на заинтересованную сторону, которое следует учесть:

a) собирались ли ПДн с целью разработки или использования *СИИ* (например, для обучения модели);

b) воздействия, вызванные нарушением конфиденциальности, связанным с обработкой ПДн, осуществляемой в целях разработки или использования *СИИ*;

c) обязательства организации перед заинтересованной стороной и ожидания от нее, связанные с обработкой ПДн, а также механизмы выполнения этих обязательств и ожиданий, включая (но не ограничиваясь):

1) справедливость;
2) риск нежелательной предвзятости, дискриминации или неточности;
3) риск профилирования и поведенческого наблюдения;
4) автоматизированное принятие решений и вмешательство человека;
5) получение информации;
6) средства для осуществления своих прав.

Оценки воздействия на защиту данных или защиту ПДн могут помочь оценить обоснованно предсказуемые воздействия на заинтересованные стороны. В *[10]* даются руководящие указания по этому вопросу. В *[11]* приводятся дополнительные требования и руководящие указания, которые помогут организации выполнить свои обязательства по защите ПДн.

6.8.2.6 Надежность

В *ГОСТ Р 71476—2024* (пункт 5.15.3) надежность — это способность системы или объекта в этой системе выполнять требующиеся от него функции в заданных условиях в течение установленного периода времени.

Для каждой идентифицированной заинтересованной стороны организация должна проанализировать обоснованно предсказуемые пользу и вред, связанные с надежностью при использовании *СИИ*.

Примеры воздействий на заинтересованную сторону, которые должны быть учтены:

- несоответствие *СИИ* заявленным требованиям к производительности на стадии эксплуатации;
- воздействия в случае, если обновление *СИИ*, дрейф данных или дрейф концепции (см. *ГОСТ Р 71476—2024*, подпункт 5.11.9.1) приведут к тому, что показатели ее производительности не будут соответствовать требованиям;
- воздействия, возникающие в результате предсказуемых сбоев, таких как ложноположительные или ложноотрицательные результаты;
- компрометация *СИИ* или ее выходных данных неуполномоченными сторонами, приводящая к невыполнению заявленных требований к производительности.

6.8.2.7 Безопасность

Под безопасностью понимается свойство *СИИ*, которое при определенных условиях предотвращает переход системы в состояние, в котором жизнь, здоровье, благополучие людей, имущество или окружающая среда находятся под угрозой. Использование *СИИ* может привести к появлению новых проблем в обеспечении безопасности. Дополнительную информацию о цели безопасности см. в *[2]* (пункт А.10).

Для каждой идентифицированной заинтересованной стороны организации следует проанализировать обоснованно предсказуемые пользу и вред, связанные с безопасностью при использовании *СИИ*.

Примеры воздействия на соответствующую заинтересованную сторону, которое следует учесть:

a) высокорисковое функционирование *СИИ*;

b) воздействия в случае изменений в *СИИ*, в том числе в результате выявления проблем или улучшения производительности *СИИ*;

c) воздействия, вызванные известными ложноположительными, ложноотрицательными результатами и другими известными сбоями.

6.8.2.8 Объяснимость

В *ГОСТ Р 71476—2024* (пункт 5.15.6) объяснимость — это свойство *СИИ* предоставлять в понятном для людей виде информацию о существенных факторах, влияющих на результаты ее функционирования. В нем также говорится, что выходные результаты некоторых *СИИ* сложнее объяснить в отличие от других из-за уровня их сложности (например, результаты глубоких нейронных сетей сложнее объяснить, чем выходные результаты деревьев решений).

Для каждой идентифицированной заинтересованной стороны организация должна проанализировать обоснованно предсказуемые пользу и вред, связанные с объяснимостью при использовании *СИИ*.

Примеры воздействия на соответствующую заинтересованную сторону, которое следует учесть:

a) доступность информации, необходимой для принятия обоснованных решений на основе выходных результатов *СИИ*, и недостающая информация имеет большое значение для конкретного варианта использования;

b) учет объяснимости на таких стадиях жизненного цикла *СИИ*, как проектирование и разработка, верификация и валидация, развертывание (например, помогает ли объяснимость заинтересованным сторонам улучшить производительность развернутой *СИИ* или принять решение о ее внедрении или отказе от него);

c) воздействие объяснимости на доверие заинтересованных сторон к *СИИ*;

d) воздействие, возникающее в результате неудачного, неправильного или неполного разъяснения.

6.8.2.9 Воздействие на окружающую среду

При оценке воздействия *СИИ* следует учитывать воздействие на окружающую среду, включая (но не ограничиваясь):

a) воздействия, возникающие на протяжении всего жизненного цикла *СИИ* от начальной стадии до вывода из эксплуатации;

b) потребление ресурсов;

c) благоприятные воздействия в результате развертывания полезных приложений ИИ, таких как защита исчезающих видов животных или оптимизация работы умных сетей электроснабжения;

d) неблагоприятные воздействия, например продвижение и стимулирование неэкологичного поведения потребителей.

Примечание — Более подробная информация об устойчивости *СИИ* приведена в *[12]*.

6.8.3 Сбои в системе искусственного интеллекта и обоснованно предсказуемое нецелевое использование

6.8.3.1 Общие положения

Заинтересованные стороны могут пострадать, если в *СИИ* случится сбой или она подвергнется обоснованно предсказуемому нецелевому использованию. При оценке воздействия *СИИ* следует учитывать:

a) прогнозируемые или известные случаи возникновения сбоев в *СИИ* и воздействие на заинтересованные стороны в результате таких потенциальных сбоев;

b) случаи обоснованно предсказуемого нецелевого использования *СИИ* и воздействие на заинтересованные стороны в результате такого использования.

6.8.3.2 Воздействие сбоя системы искусственного интеллекта

Организации следует документировать потенциальные сбои *СИИ*. К ним могут относиться потенциальные типы случайных входных данных в *СИИ*, а также сбои отдельных компонентов или процессов *СИИ*. К ним также можно отнести человеческий фактор, сбои в работе аппаратного обеспечения или ПО, а также нецелевое взаимодействие систем или процессов.

Организация должна определить воздействия на выявленные заинтересованные стороны (см. 6.7) в результате сбоев в *СИИ*. Последствия и воздействие на конкретные группы лиц или социальные группы из-за сбоев в *СИИ* должны быть задокументированы.

6.8.3.3 Влияние обоснованно предсказуемого нецелевого использования системы искусственного интеллекта

Организация должна документировать обоснованно предсказуемое нецелевое использование *СИИ*. Это может быть полезно для оценки выгоды злоумышленников, намеренно использующих *СИИ* неправильно.

Организация должна определить и задокументировать обоснованно предсказуемые воздействия на выявленные заинтересованные стороны (см. 6.7) в результате типичного обоснованно предсказуемого нецелевого использования *СИИ*.

6.9 Меры по устранению вреда и обеспечению пользы

Организация должна рассмотреть возможность документирования мер как на техническом, так и на управленческом уровнях для устранения вреда и обеспечения пользы в рамках оценки воздействия *СИИ*, например:

a) предложения по внедрению мер на техническом уровне. Эти предложения следует учитывать в рамках конкретных процессов утверждения *СИИ* (см. 5.11).

Примеры

1 Команда разработчиков предлагает добавить шифрование в реестр пользователей после выявления обоснованно предсказуемых неблагоприятных воздействий на защиту ПДн.

2 Стратегическое управление рекомендуется для обеспечения того, чтобы разработка или использование СИИ осуществлялись надлежащим образом и чтобы воздействия прорабатывались в течение всего жизненного цикла СИИ.

3 Операторы СИИ могут принять решение о выводе из эксплуатации, отзыве или обновлении СИИ, если ее показатели производительности не соответствуют требованиям. Вывод СИИ из эксплуатации может оказать воздействие на заинтересованные стороны (например, вызвать финансовые потери или риск для здоровья);

b) рекомендации для рассмотрения высшим руководством при разработке и внедрении политик организации в отношении *СИИ*.

Примеры

1 Команда разработчиков выявляет потенциальные возможности использования системы генерации текстов потенциальным клиентом, который использует систему для создания коммерческих предложений, которые могут быть ошибочно ассоциированы с физическими лицами. Вопрос о продаже системы такому клиенту относится к сфере деловых отношений и может быть учтен при заключении контракта.

2 Обучение и повышение осведомленности рекомендуется не только для пользователей СИИ, но и для других сотрудников организации, а также иных заинтересованных сторон, например, о воздействии, связанном с автоматизированным принятием решений. Вред может возникнуть из-за отсутствия или недостаточного уровня осведомленности и подготовки.

3 Организация может установить свои приоритеты для оценки и создать инфраструктуру для отображения мер по реализации этих приоритетов. К ним могут относиться защита ПДн, трудовые отношения, нежелательная предвзятость, дискриминация, прозрачность, подотчетность и другие принципы.

Указанные рекомендации следует учитывать при подготовке и обновлении плана по регулированию рисков.

Приложение А
(справочное)

Руководящие указания по использованию настоящего стандарта совместно с *ГОСТ Р ИСО/МЭК 42001*

В *ГОСТ Р ИСО/МЭК 42001* содержатся требования и рекомендации по оценке воздействия *СИИ*.

В таблице А.1 приведены указания по использованию настоящего стандарта в дополнение к *ГОСТ Р ИСО/МЭК 42001*.

Таблица А.1 — Структурные элементы *ГОСТ Р ИСО/МЭК 42001*, относящиеся к настоящему стандарту

ГОСТ Р ИСО/МЭК 42001—2024	Тип	Краткое содержание	Настоящий стандарт	Пояснение, как настоящий стандарт соотносится с *ГОСТ Р ИСО/МЭК 42001*
6.1.4 Оценка воздействия *СИИ*	Требование	В *ГОСТ Р ИСО/МЭК 42001—2024* (пункт 6.1.4) рассматривается оценка воздействия *СИИ* в целом: организации требуется создать процесс оценки воздействия для определения того, какое воздействие *СИИ* может оказать на отдельных лиц и социальные группы. Кроме того, в нем рассматривается необходимость документирования. Данный пункт рассматривают на уровне системы управления, а также в нем описывается необходимость целостного рассмотрения оценок воздействия системы и их взаимосвязь с организационным риском	Все положения стандарта	В настоящем стандарте содержатся руководящие указания по выполнению требований *ГОСТ Р ИСО/МЭК 42001—2024* (пункт 6.1.4). В разделе 5 описаны процессы, которые должны быть внедрены организацией, а в разделе 6 содержится подробная информация о документации. Настоящий стандарт может рассматриваться в первую очередь как помощь организации в установлении процессов и создании документации для оценок воздействий на уровне отдельной системы. Однако такие отдельные оценки допускается рассматривать в комплексе как необходимые исходные данные для *ГОСТ Р ИСО/МЭК 42001—2024* (пункт 6.1.4)
8.4 Оценка воздействия *СИИ*	Требование	В *ГОСТ Р ИСО/МЭК 42001—2024* (подраздел 8.4) требуется, чтобы оценка влияния *СИИ* проводилась через запланированные промежутки времени или в случае существенных изменений. Данный подраздел рассматривают на уровне системы управления, а также в нем описывается необходимость целостного рассмотрения оценок воздействия *СИИ* и их взаимосвязь с риском организации	5.4	В 5.4 настоящего стандарта описываются сроки и возможные поводы для оценки воздействия *СИИ*. Настоящий стандарт может рассматриваться в первую очередь как помощь организации в установлении процессов и создании документации для оценки воздействий на уровне отдельных систем. Однако такие отдельные оценки допускается рассматривать в комплексе как необходимые исходные данные для *ГОСТ Р ИСО/МЭК 42001—2024* (подраздел 8.4)

Продолжение таблицы А.1

ГОСТ Р ИСО/МЭК 42001—2024	Тип	Краткое содержание	Настоящий стандарт	Пояснение, как настоящий стандарт соотносится с *ГОСТ Р ИСО/МЭК 42001*
В.2.2 Политика в области ИИ	Руководство по внедрению	В руководстве по внедрению говорится, что необходимость проведения оценки воздействия *СИИ* следует задокументировать в политике в сфере ИИ	Все положения стандарта	Настоящий стандарт может быть использован организацией для поддержки своей политики в сфере ИИ. В разделе 5 говорится о создании и внедрении процесса оценки воздействия *СИИ*, который может быть полезен при создании документации по политике в сфере ИИ
В.3.2 Роли и обязанности в области ИИ	Руководство по внедрению	В руководстве по внедрению говорится, что организация должна учитывать оценку воздействия *СИИ* при распределении ролей и обязанностей	5.5	В настоящем разделе описаны дополнительные факторы, которые необходимо учитывать при определении ролей и обязанностей для оценки воздействия *СИИ*
В.4.2 Документация ресурсов	Руководство по внедрению	В руководстве по внедрению перечислены следующие ресурсы: - компоненты *СИИ*; - ресурсы для данных; - инструментальные ресурсы; - системные и вычислительные ресурсы; - человеческие ресурсы	6.6	В 6.6 настоящего стандарта рассматриваются различные аспекты, упомянутые в *ГОСТ Р ИСО/МЭК 42001—2024* (пункт В.4.2), поскольку документирование этих элементов имеет решающее значение для возможности оценки воздействий *СИИ*. Возможно, в некоторых случаях документирование ресурсов в рамках оценки воздействия может удовлетворить общую потребность в документировании ресурсов; организация должна решить этот вопрос самостоятельно
В.5.1 Цели	Цель контроля	Согласно данной цели, воздействия *СИИ* на отдельных лиц и социальные группы, подверженные влиянию *СИИ*, следует оценивать на протяжении всего ее жизненного цикла	Все положения стандарта	В настоящем стандарте содержатся руководящие указания по достижению этой цели
В.5.2 Процесс оценки воздействия *СИИ*	Руководящие указания по контролю и внедрению	Контроль касается процесса оценки воздействия *СИИ*	Раздел 5	В разделе 5 настоящего стандарта содержатся руководящие указания по созданию и внедрению процессов оценки воздействия *СИИ* с предоставлением подробных руководящих указаний по осуществлению данного контроля организацией

Окончание таблицы А.1

ГОСТ Р ИСО/МЭК 42001—2024	Тип	Краткое содержание	Настоящий стандарт	Пояснение, как настоящий стандарт соотносится с *ГОСТ Р ИСО/МЭК 42001*
B.5.3 Документация по оценке воздействия *СИИ*	Руководящие указания по контролю и внедрению	Контроль касается документирования оценки воздействия *СИИ*	Раздел 6	В разделе 6 настоящего стандарта содержатся подробные руководящие указания по документированию организацией оценки воздействия *СИИ*
B.5.4 Оценка воздействия *СИИ* на отдельных лиц и группы лиц	Руководящие указания по контролю и внедрению	Контроль касается оценки обоснованно предсказуемых воздействий на отдельных лиц и группы лиц	6.7	Различные аспекты, упомянутые в *ГОСТ Р ИСО/МЭК 42001—2024* (пункт B.5.4), более подробно рассматриваются в 6.7 настоящего стандарта
B.5.5 Оценка воздействия *СИИ* на общество	Руководящие указания по контролю и внедрению	Контроль касается оценки обоснованно предсказуемых воздействий на общество	6.7	Различные аспекты, упомянутые в *ГОСТ Р ИСО/МЭК 42001—2024* (пункт B.5.5), более подробно рассматриваются в 6.7 настоящего стандарта
B.6.1.3 Процессы ответственного проектирования и разработки *СИИ*	Руководство по внедрению	Контроль касается внедрения и документирования процессов оценки воздействия *СИИ*	Разделы 5, 6	В обоих разделах настоящего стандарта рассматривается интеграция оценки воздействия в общий жизненный цикл ИИ

Приложение В
(справочное)

Руководящие указания по использованию настоящего стандарта совместно с *[2]*

В.1 Общие положения

В *[2]* содержатся руководящие указания по управлению рисками для организаций, разрабатывающих, предоставляющих или использующих *СИИ*. Стандарт *[2]* основан на *ГОСТ Р ИСО 31000*, в котором описывается управление организационными рисками в целом. В *[2]* объясняется, как принципы, структуры и процессы *ГОСТ Р ИСО 31000* могут применяться в контексте ИИ. В настоящем приложении содержатся руководящие указания по поддержке управления рисками с помощью оценки воздействия *СИИ*.

В.2 Различия между управлением рисками и оценкой воздействия системы искусственного интеллекта

В *ГОСТ Р ИСО 31000* (и, соответственно, в *[2]*) управление рисками описывается как «скоординированные действия для управления организацией в отношении рисков» (см. *ГОСТ Р ИСО 31000—2019*, пункт 3.2). Менеджмент риска «является итеративным процессом и помогает организациям в определении стратегии, достижении целей и принятии обоснованных решений» (см. *ГОСТ Р ИСО 31000—2019*, введение). Таким образом, управление рисками — это общая управленческая деятельность, которая затрагивает всю организацию в целом, включая то, как разработчики *СИИ* ведут проекты по разработке, поставщики *СИИ* управляют отношениями со своими клиентами, а пользователи *СИИ* используют функции *СИИ*. Оценка воздействия *СИИ*, в отличие от общего управления рисками, направлена на обоснованно предсказуемые воздействия использования *СИИ*:

a) на ограниченный круг соответствующих заинтересованных сторон, а именно на отдельных лиц, группы лиц, социальные группы и окружающую среду;

b) на потенциальное или конкретное использование *СИИ* в отличие от общих вопросов стратегического управления и руководства, таких как бизнес-стратегии, управление соответствием, финансовое управление или технологические стратегии.

Таким образом, оценка воздействия *СИИ* в большей степени, по сравнению с управлением рисками, ориентируется на продукт или услугу. Оценка также предназначена для непосредственного выполнения командами, занимающимися разработкой, поставкой или техническим управлением *СИИ*.

Оценки воздействия *СИИ* являются важной частью общего цикла управления рисками и неотъемлемым вкладом в успешную оценку рисков.

В.3 Принципы управления рисками по отношению к оценке воздействия системы искусственного интеллекта

На рисунке В.1 представлена взаимосвязь *[2]* с настоящим стандартом. Для наглядности на рисунке В.1 для различных частей настоящего стандарта указаны соответствующие подразделы *[2]*.

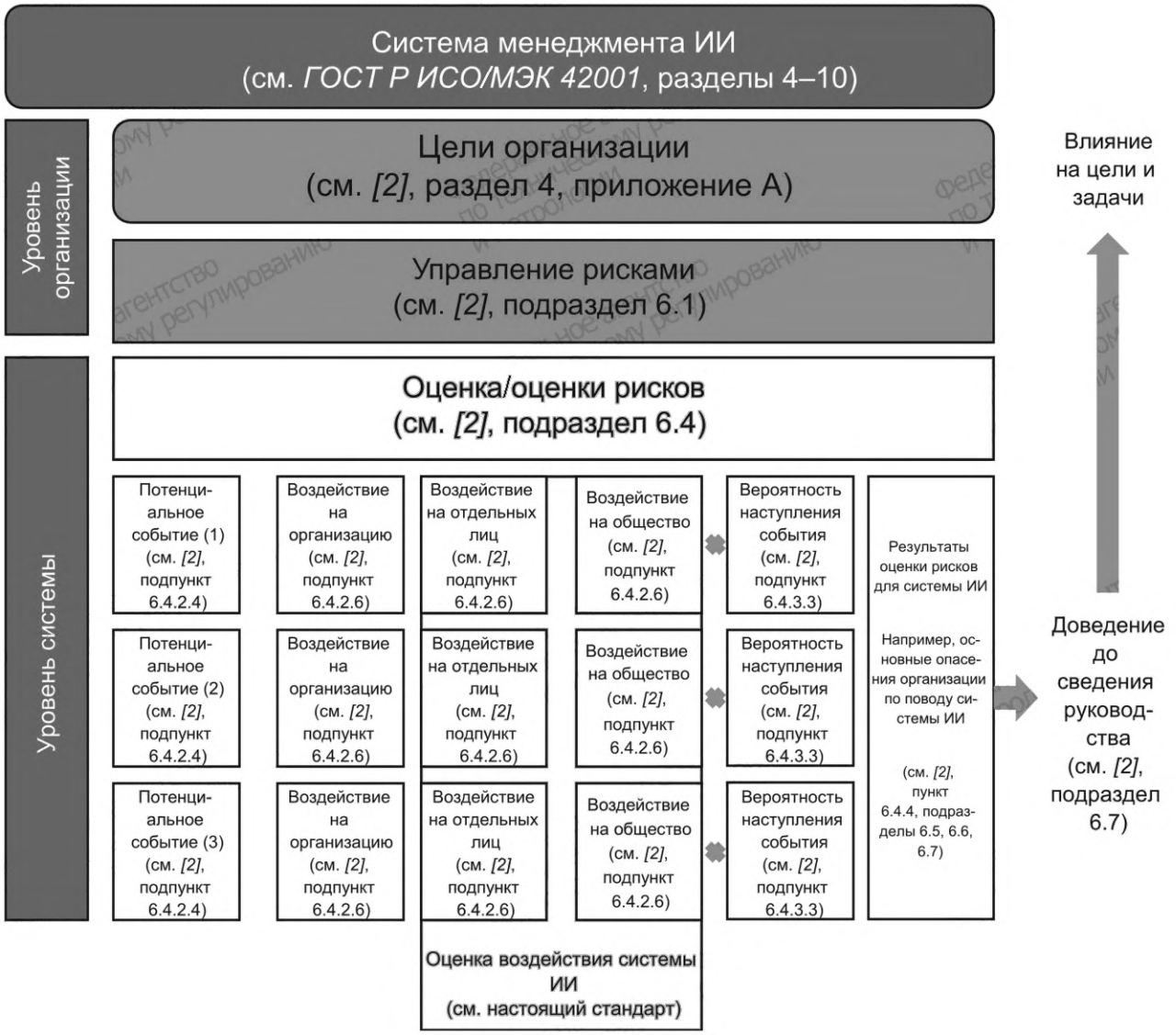

Рисунок В.1 — Взаимосвязь *[2]* с настоящим стандартом

Приложение С
(справочное)

Таксономия вреда и пользы

Таблица С.1 представляет собой шаблон для организации по анализу обоснованно предсказуемых вреда и пользы *СИИ*. Организация должна учитывать руководящие указания для каждого типа заинтересованных сторон, имеющих отношение к *СИИ*, которые подлежат оценке воздействия *СИИ*.

Таблица С.1 — Вред и польза

Цели	Пример вреда	Пример пользы
Подотчетность		
Надзор за значительными неблагоприятными воздействиями	Угроза возможностям и правовому статусу. Угроза человеческой жизни. Психологический или физический вред	Значительные неблагоприятные воздействия удается предотвратить до того, как они успеют нанести вред
Соответствие цели	*СИИ* выдает неверные результаты и данные	Чем лучше *СИИ* решает поставленную перед ней задачу, тем больше пользы для заинтересованных сторон
Стратегическое управление данными	Отсутствие должного управления данными (разметки, проставления тегов, утверждения, оценки соответствия и т. д.) может привести к несоответствию назначению и нежелательной предвзятости, а также к другим проблемам	Надлежащее управление данными позволяет выявить потенциальные проблемы (нежелательная предвзятость, несоответствие назначению и т. д.) еще до обучения моделей, что позволяет добиться лучших результатов для соответствующих заинтересованных сторон
Человеческий надзор и контроль	Отсутствие надзора и контроля со стороны человека может привести к затягиванию устранения проблем и угроз	Надлежащий надзор и контроль позволяет выявлять проблемы на ранней стадии и устранять их. Надзор и контроль со стороны человека могут способствовать открытию новых знаний и усовершенствованию *СИИ*, что принесет пользу соответствующим заинтересованным сторонам
Прозрачность и объяснимость		
Объяснимость *СИИ*	Соответствующие заинтересованные стороны неверно истолковывают или неправильно понимают предполагаемое назначение или использование *СИИ* на основе полученных результатов. Соответствующие заинтересованные стороны не могут полагаться на *СИИ* в принятии обоснованных решений. Чрезмерное доверие автоматизации	Соответствующий уровень объяснимости может способствовать выстраиванию доверия и взаимопонимания между соответствующими заинтересованными сторонами. Признаки объяснимости могут помочь отследить ошибки в процессе принятия решений *СИИ*
Обмен информацией с соответствующими заинтересованными сторонами	Соответствующие заинтересованные стороны не знают об ограничениях, параметрах и настройках *СИИ*, которые могут обеспечить ее ответственное использование. Критически важная информация по безопасности не является легкодоступной или понятной	Соответствующие заинтересованные стороны могут быть проинформированы о целевом использовании, возможностях и ограничениях *СИИ*, что снижает риск недопонимания или нецелевого использования. *СИИ* может безопасно использоваться соответствующими заинтересованными сторонами, а в случае с поставщиками и третьими сторонами могут быть выполнены обязательства по безопасности

Окончание таблицы С.1

Цели	Пример вреда	Пример пользы
Раскрытие информации о взаимодействии с ИИ	Соответствующие заинтересованные стороны могут взаимодействовать с *СИИ*, полагая, что это человек	Соответствующие заинтересованные стороны могут принимать обоснованные решения и давать согласие на взаимодействие с *СИИ*, а не с человеком, включая разглашение конфиденциальной информации
Справедливость		
Качество услуг	Качество услуг неодинаково для разных демографических групп, что приводит к неравенству (более высокие показатели для одних групп по сравнению с другими)	Качество обслуживания соответствует демографическим показателям, открывая новые возможности для исторически менее привилегированных групп
Распределение возможностей	Результаты *СИИ* приводят к неравномерному распределению возможностей среди различных демографических групп	Результаты *СИИ* позволяют выявить и смягчить неравномерное распределение возможностей для различных демографических групп
Минимизация стереотипов	Результаты *СИИ* приводят к формированию стереотипов, стиранию или уничижению демографических групп	Результаты *СИИ* помогают выявить тенденции в формировании стереотипов, стирании или унижении демографических групп, что позволяет найти решения данных проблем
Надежность		
Неудачи и меры по их устранению	Сбои (как предсказуемые, так и неизвестные) не имеют планов по устранению, и *СИИ*, предоставляющая критически важные услуги, не может функционировать	Хорошо спланированная система с надежными планами резервного копирования, устранения неисправностей и устранения последствий может способствовать повышению надежности системы
Мониторинг, обратная связь, оценка	Отсутствие мониторинга приводит к тому, что наборы данных устаревают, а результаты работы *СИИ* становятся ненадежными	Оценка систем находит ошибки до начала применения
Безопасность и защита ПДн		
Защита ПДн и соблюдение правовых обязательств и политик	Данные, используемые для обучения моделей, не защищены должным образом с точки зрения защиты ПДн, и ПДн используются в системах, которые не должны их использовать	*СИИ* обучаются на соответствующим образом деидентифицированных ПДн, гарантируя, что система может приносить пользу, которая извлекается из ПДн (например, данных о здоровье) без риска для отдельных лиц
Безопасность системы и соблюдение правовых обязательств и политик	Уязвимости в безопасности системы могут позволить злоумышленникам захватить контроль над системой и применять ее по непредусмотренному назначению	Постоянно обучающиеся системы могут прогнозировать новые и зарождающиеся угрозы безопасности и защищаться от них

Приложение D
(справочное)

Совмещение оценки воздействия системы искусственного интеллекта с другими оценками

D.1 Общие положения

Цель настоящего приложения — предоставить руководящие указания по внедрению оценки воздействия *СИИ* в сочетании с существующими или смежными оценками таким образом, чтобы отразить требования *ГОСТ Р ИСО/МЭК 42001* и существующих требований к оценке (например, рассмотрение рисков, правовых вопросов, вопросов безопасности, конфиденциальности, этики данных, прав человека, закупок и архитектуры). Цель состоит в том, чтобы облегчить эти оценки и избежать их дублирования.

При эффективном и надлежащем проведении оценка воздействия *СИИ* может дополнять и использовать другие существующие проверки, причем потенциально многократно в течение жизненного цикла *СИИ*.

Данное приложение состоит из трех частей, которые могут помочь организации при проведении оценки воздействия ИИ:

- D.2 «Руководство по координации»: Руководство, помогающее организации координировать и развивать существующие оценки воздействия, чтобы более эффективно и результативно выполнять требования *ГОСТ Р ИСО/МЭК 42001* с помощью существующих процессов оценки воздействия, а также избежать дублирования;
- D.3 «Руководство по согласованию оценки воздействия»: Руководство, которое поможет организации определить, как существующие оценки воздействия соотносятся с оценкой воздействия *СИИ* и могут быть использованы для ускорения процесса завершения оценки воздействия *СИИ*;
- D.4 «Руководство по сопоставлению»: Наглядное руководство, помогающее сопоставить предлагаемую оценку воздействия *СИИ* с существующими в организации или смежными процессами оценки воздействия. В настоящем руководстве поясняются возможности для координации, уменьшается дублирование, чтобы оценка воздействия *СИИ* была более эффективно встроена в общую систему управления в организации.

D.2 Руководство по координации

Многие организации уже регулярно проводят различные виды оценки воздействия ИТ-систем (например, оценки рисков, конфиденциальности и прав человека). Несмотря на то, что каждая оценка проводится с определенной целью, у них часто имеются сходства как в отношении соответствующих заинтересованных сторон, так и в отношении требуемых данных. Добавление оценки воздействия конкретно для *СИИ* без учета того, как она может использовать другие потенциально требуемые оценки и вносить вклад в их реализацию, может привести к ненужным разногласиям, затратам и задержкам в ответственном использовании *СИИ*.

Один из подходов, который организация может применять, чтобы ускорить оценки воздействия и избежать дублирования процессов, заключается в развертывании руководства по координации, которое помогает согласовать соответствующие проверки, необходимые для оценки воздействия *СИИ* (например, проверка рисков, правовая проверка, проверка безопасности, конфиденциальности, соблюдения этики данных, прав человека, проверка закупок или архитектуры). Чтобы обеспечить эффективность руководства по координации, организация может рассмотреть возможность назначения координатора или команды, обладающих опытом в одной или нескольких перечисленных областях, которые будут выступать в качестве координаторов соответствующих внутренних функций.

При внедрении руководства по координации необходимо учитывать следующие ключевые аспекты:

a) определение того, какие *СИИ* будут проверяться (например, которые входят в область применения системы менеджмента ИИ) и как другие проведенные или планируемые к проведению организацией оценки воздействия могут быть использованы при проведении такого анализа.

Примечание — Более подробные руководящие указания по области применения оценки воздействия *СИИ* приведены в *ГОСТ Р ИСО/МЭК 42001—2024* (подраздел 5.1);

b) в оценки воздействия *СИИ* различных типов может вовлекаться целый ряд внешних соответствующих заинтересованных сторон. Целостный и скоординированный подход к привлечению соответствующих заинтересованных сторон помогает организации понять совокупный риск и риск на уровне системы и управлять им, что, в свою очередь, повышает полезность любой отдельной, индивидуальной оценки;

c) многие аспекты, рассматриваемые в различных оценках воздействия, такие как кибербезопасность и конфиденциальность, одинаковы, но подход к проверке немного отличается. Руководство по координации может помочь определить, какие основания, вопросы и данные, как правило, схожи и, следовательно, повторяются в оценках воздействия разных типов, а также где существуют критические различия, которые можно понять и выстроить по их приоритетности (например, кибербезопасность и конфиденциальность);

d) одновременное проведение проверок может способствовать более эффективному использованию ресурсов и более быстрому выполнению проверок. Например, оценка воздействия на конфиденциальность может включать информацию о данных и их качестве. Если такая оценка проводится параллельно с проверкой *СИИ*, можно

одновременно рассмотреть и устранить воздействия, связанные с информацией о данных и их качеством в обоих контекстах;

е) установление компонентов *СИИ* может определить ряд соответствующих оценок воздействия, которые могут внести вклад в оценку воздействия *СИИ*. Это может помочь понять, как множественные *СИИ* в организации могут иметь зависимости, порождающие неожиданное поведение и воздействия. Один из подходов заключается в группировке *СИИ* по вариантам использования. Например, все системы, используемые для кредитования, допускается сгруппировать в рамках одной оценки. В свою очередь, эта оценка может обновляться при каждом изменении *СИИ*, что позволяет организации оценить воздействие изменений в *СИИ* и избежать дублирования. Когда составные компоненты *СИИ* идентифицированы, организация может лучше определить ряд оценок воздействия, соответствующих оценке воздействия *СИИ*, что позволит улучшить координацию между этими оценками.

D.3 Руководство по согласованию оценки воздействия

Существует множество оценок воздействия, которые организации, как правило, проводят в рамках выполнения существующих требований к оценке. Существующие оценки, проведенные организацией, могут содержать ценные материалы для оценки воздействия *СИИ* в соответствии с данным стандартом.

Пример процесса разработки руководства по согласованию оценки воздействия:

- шаг 1. Идентифицировать существующие оценки воздействия, которые могут соответствовать проводимой оценке воздействия ИИ. Как правило, к существующим оценкам воздействия, которые могут соответствовать проводимой оценке воздействия ИИ, относятся:
  - любые оценки воздействия на права человека (HRIA), относящиеся к деятельности организации, которая будет каким-либо образом дополнена или затронута оцениваемым ИИ;
  - любые оценки воздействия на конфиденциальность (PIA), относящиеся к данным, которые будут обрабатываться в *СИИ*;
  - любые оценки воздействия на окружающую среду (EIA), относящиеся к среде, на которую будет воздействовать или в которой будет работать *СИИ*;
  - любые оценки воздействия на безопасность (SIA), относящиеся к изменениям в информационной системе организации, связанным с *СИИ*;
  - любые оценки финансового воздействия (FIA), относящиеся к расходам и доходам, связанным с *СИИ*;
  - любые оценки воздействия на деятельность (BIA), относящиеся к потенциальным последствиям для бизнес-функций и процессов, связанных с *СИИ*;
- шаг 2. Извлечь из существующих оценок воздействия соответствующее содержание, которое может соответствовать проводимой оценке воздействия ИИ, как показано в таблице D.1.

Таблица D.1 представляет собой матрицу, демонстрирующую, как организация может идентифицировать согласование между существующей оценкой воздействия и оценкой воздействия *СИИ*. Руководство по согласованию оценок воздействия может использоваться для обозначения того, какие разделы и подразделы оценки воздействия *СИИ*, описанные в настоящем стандарте, должны быть основаны на существующих оценках воздействия. Таблица D.1 приводится только в качестве примера и может быть изменена в соответствии с потребностями организации.

Таблица D.1 — Наглядный пример руководства по согласованию оценки воздействия

Ряд	Подразделы настоящего стандарта	Пункты настоящего стандарта	Потенциальный вклад соответствующей HRIA	Потенциальный вклад соответствующей PIA	Потенциальный вклад соответствующей EIA	Потенциальный вклад соответствующей FIA	Потенциальный вклад соответствующей BIA	Потенциальный вклад соответствующей SIA
1	5.1	Н/Д	X	X	(заполняется организацией)			
2	5.2	Н/Д	X	X				
3	5.3	Н/Д	X	X				
4	5.4	5.4.1	—	—				
5		5.4.2	—	X				
6		5.4.3	—	X				
7		5.4.4	—	X				
8	5.5	Н/Д	X	X				
9	5.6	Н/Д	X	X				
10	5.7	Н/Д	X	X				
11	5.8	Н/Д	X	X				
Примечание — Вклад, который может относиться к оценке воздействия *СИИ*.								

D.4 Руководство по сопоставлению

Руководство по сопоставлению — это наглядное руководство с пояснительным материалом для демонстрации, как оценка воздействия *СИИ* может использоваться в сочетании с другими существующими или смежными оценками для того, чтобы:

- избежать повторений и дублирования; и
- комплексно изучить и выявить источники риска *СИИ*.

Руководство по сопоставлению — это просто концептуальная отправная точка, помогающая организации применять комплексный подход к оценке воздействия. Этот инструмент призван проиллюстрировать:

- ряд существующих оценок воздействия, которые связаны или пересекаются с оценками воздействия *СИИ*;
- общий контекст у различных оценок воздействия (например, цели организации, задачи управления, риски или лидерство), при этом допускаются различия в предметных областях (например, ИИ, конфиденциальность, качество или безопасность).

Целью является поддержка и отражение стандартного подхода системы менеджмента, где гармонизированная структура обеспечивает общность, допуская при этом наличие вариативности.

На рисунке D.1 показана потенциальная взаимосвязь между оценкой воздействия *СИИ* и другими оценками воздействия в различных областях в рамках более широкой организации. Рисунок D.1 демонстрирует один из возможных примеров внедрения.



Рисунок D.1 — Потенциальная взаимосвязь между оценкой воздействия системы искусственного интеллекта и другими оценками

Приложение Е
(справочное)

Пример шаблона оценки воздействия системы искусственного интеллекта

Е.1 Общие положения

В настоящем приложении приводится пример шаблона, который организация может применить и адаптировать для документирования оценки воздействия *СИИ*. Поля в шаблоне основаны на подразделах в разделе 6. Предполагается, что для каждой разрабатываемой или используемой организацией *СИИ* проведена как минимум одна оценка воздействия *СИИ*. Однако может возникнуть необходимость в проведении нескольких оценок воздействия *СИИ* для какой-либо системы в течение ее жизненного цикла.

Примечания

1 Вероятно, политика организации в области ИИ содержит требования по выполнению оценки воздействия *СИИ*. В некоторых случаях политика организации в области ИИ может предписывать проведение оценки воздействия *СИИ* для каждой *СИИ*, которую организация разрабатывает или использует. В других случаях политика организации в области ИИ может включать условия, для выполнения которых требуется проведение оценки воздействия *СИИ*.

2 В примере, приведенном в Е.2, показана самостоятельная оценка воздействия *СИИ*. Однако оценку воздействия *СИИ* допускается включать в существующие оценки воздействия.

Е.2 Пример шаблона оценки воздействия искусственного интеллекта

Е.2.1 Оценка воздействия системы искусственного интеллекта для [название или идентификация системы]

В ходе оценки воздействия *СИИ* идентифицируются обоснованно предсказуемые воздействия *СИИ* на отдельных лиц, группы лиц и социальные группы; устанавливается, является ли предусмотренное использование *СИИ* высокорисковым или ограниченным и требуется ли проведение дополнительных проверок и утверждений. Заполненная документация по оценке воздействия *СИИ*, включая настоящий шаблон, требует утверждения [имя утверждающего лица или наименование ответственного подразделения].

Е.2.2 Системная информация

Е.2.2.1 Общие сведения

Идентификация *СИИ*, являющейся объектом данной оценки воздействия *СИИ* (см. таблицу Е.1).

Таблица Е.1 — Общая информация о *СИИ*

Название или идентификация *СИИ*	Информация для идентификации оцениваемой *СИИ*
Другие идентификаторы *СИИ*	Альтернативные средства идентификации оцениваемой *СИИ*
Стадия жизненного цикла *СИИ*	Информация о стадии жизненного цикла, на которой проводится оценка воздействия *СИИ*

Примечание — В последнем поле в таблице Е.1, касающемся идентификации стадии жизненного цикла *СИИ*, указывается, на какой стадии жизненного цикла *СИИ* проводится данная оценка воздействия *СИИ*. В нем также указывается, на какой стадии жизненного цикла продукта проводится оценка воздействия *СИИ*, если *СИИ* разрабатывается организацией.

Отслеживание истории пересмотра данной оценки воздействия *СИИ* (см. таблицу Е.2).

Таблица Е.2 — История пересмотра

Автор	Лицо или команда, ответственные за оценку воздействия *СИИ*
Дата последнего изменения	Дата последнего изменения оценки воздействия *СИИ* (при необходимости можно добавить дополнительные строки)

Сведения об экспертах и утверждающих лицах данной оценки воздействия *СИИ* (см. таблицу Е.3).

Таблица Е.3 — Проверка и утверждение

Проверено	Лицо или команда, ответственные за проверку оценки воздействия *СИИ*
Дата проверки	Дата проведения проверки
Утверждено	Лицо или команда, ответственные за утверждение оценки воздействия *СИИ*
Дата утверждения	Дата утверждения проверки

Е.2.2.2 Описание системы искусственного интеллекта

Включает в себя базовое описание *СИИ*, ее деятельности и высокоуровневое описание принципа ее работы. Допускается включить любые приложения, которые могут помочь рецензентам лучше понять *СИИ* и среду, в которой она работает (см. таблицу Е.4 и 6.3.1).

Таблица Е.4 — Базовое описание *СИИ*

Базовое описание *СИИ* согласно описанию в 6.3.1

Е.2.2.3 Функции и возможности системы искусственного интеллекта

Включает подробное описание функций и возможностей *СИИ*, а также информацию о том, планируются они или уже существуют (см. таблицу Е.5). Включает все зависимости от других систем (см. 6.3.2).

Таблица Е.5 — Описание функций и возможностей *СИИ*

Номер	Функции или возможности *СИИ*	В текущей версии?	Планируется в будущем или уже запланировано к внедрению?
1	Описание функций или возможностей согласно описанию в 6.3.2	«Да» или «нет»	В запланированной версии? Предполагаемая версия и дата
2	Описание функций или возможностей, как описано в 6.3.2	«Да» или «нет»	В запланированной версии? Предполагаемая версия и дата
3	…	…	…

Подробные сведения о зависимостях от других систем (см. таблицу Е.6).

Таблица Е.6 — Описание зависимостей *СИИ* от других систем

Информация о потенциальных зависимостях оцениваемой *СИИ* от других систем, их функций и возможностей

Е.2.2.4 Назначение системы искусственного интеллекта

Включает описание назначения *СИИ*, причин ее разработки или применения, ценностного предложения *СИИ* и ожидаемых результатов от использования *СИИ* (см. таблицу Е.7). См. 6.3.3.

Таблица Е.7 — Описание назначения *СИИ*

Информация о назначении *СИИ* согласно описанию в 6.3.3

Е.2.2.5 Целевое использование

Включает любое предполагаемое целевое использование *СИИ*, для которого она была разработана (см. таблицу Е.8 и 6.3.4).

Таблица Е.8 — Описание целевого использования СИИ

Номер	Целевое использование	Описание целевого использования. Включает информацию о предполагаемом конечном пользователе *СИИ*, о том, где и когда *СИИ* может быть использована
1	Выявление целевого использования	Описание целевого использования *СИИ* согласно описанию в 6.3.4
2	Выявление целевого использования	Описание целевого использования *СИИ* согласно описанию в 6.3.4
3	…	…

Е.2.2.6 Нецелевое использование

Содержит описание обоснованно предсказуемого нецелевого использования (см. таблицу Е.9 и 6.3.5).

Таблица Е.9 — Описание нецелевого использования *СИИ*

Номер	Нецелевое использование	Описание нецелевого использования. Следует указать, является ли такое нецелевое использование обоснованно предсказуемым нецелевым использованием *СИИ*
1	Выявление нецелевого использования	Описание нецелевого использования *СИИ* согласно описанию в 6.3.5
2	Выявление нецелевого использования	Описание нецелевого использования *СИИ* согласно описанию в 6.3.5
3	…	…

Е.2.3 Информация о данных и их качество

Е.2.3.1 Информация о данных

Содержит информацию о наборах данных, используемых в *СИИ* (см. 6.4.2).

Для каждого набора данных можно заполнить таблицу Е.10.

Таблица Е.10 — Информация о данных

Название набора данных, версия, размер	Идентификация набора данных и дополнительная информация при необходимости
Владелец набора данных	Владелец набора данных, если применимо
Права доступа к набору данных	Информация, связанная с контролем доступа
Описание содержания набора данных	Может включать информацию о временном охвате данных, о том, являются ли наборы данных реальными, синтетическими и полусинтетическими, о географических регионах, охватываемых набором данных, и т. д.
Предусмотренные и непредусмотренные назначения этого набора данных, в том числе утверждено ли его использование для *СИИ*	В соответствии с указаниями
Известные или обоснованно предсказуемые риски нежелательной предвзятости в наборе данных и охватываемых географических регионах	В соответствии с указаниями
Процессы, применяемые к набору данных для обеспечения качества	Описание процессов обеспечения качества данных применительно к набору данных
Необходима ли оценка воздействия на защиту данных?	«Да» или «нет», в зависимости от наличия ПДн в наборе данных

Е.2.3.2 Документирование качества данных

Содержит информацию о характеристиках качества данных, которые учитывались для каждого набора данных, используемого в *СИИ* (см. таблицу Е.11 и 6.4.3).

Таблица Е.11 — Информация о качестве данных

Удовлетворенные характеристики качества данных		
Номер	Характеристика качества данных	Значение для *СИИ*, дата выполнения, соответствующие утверждения
1	Название и описание характеристики качества данных	Информация о качестве данных согласно описанию в 6.4.3
2	Название и описание характеристики качества данных	Информация о качестве данных согласно описанию в 6.4.3
3	…	…
Запланированные, но еще не выполненные характеристики качества данных. Дата, к которой планируется выполнить характеристику качества данных.		
Номер	Характеристика качества данных	Значимость для системы, планируемый срок удовлетворения
1	Название и описание характеристики качества данных	Информация о качестве данных согласно описанию в 6.4.3
2	Название и описание характеристики качества данных	Информация о качестве данных согласно описанию в 6.4.3
3	…	…

Е.2.4 Информация об алгоритмах и моделях

Е.2.4.1 Информация об алгоритмах

Включает информацию о каждом алгоритме, разработанном для *СИИ* или используемом в ней (см. таблицу Е.12 и 6.5.2, 6.5.3).

Таблица Е.12 — Информация об алгоритме

Пригодность алгоритма для целей *СИИ* и ее целевого использования		Соответствует целям деятельности и выполняемой задачи
Происхождение алгоритма		Пример: научные статьи, журналы и презентации или репозитории исходного кода
Валидность алгоритма		Пример: проверен и доказан авторитетными экспертами в данной области
Использование в реальной практике и показатели деятельности		Информация о предыдущем использовании алгоритма согласно описанию в 6.5.2, перечисление d)
Требования к данным		В соответствии с указаниями
Подверженность нежелательным последствиям		Пример: ложные корреляции, чрезмерная подгонка или нежелательная предвзятость
Юридические, организационные или технические требования		
1	Название и описание требования	Описание того, как было выполнено требование
2	Название и описание требования	Описание того, как было выполнено требование
3	…	…
Невыполненные юридические, организационные или технические требования		
1	Название и описание требования	Планы по выполнению требования
2	Название и описание требования	Планы по выполнению требования
3	…	…

Е.2.4.2 Информация о моделях, используемых в системе искусственного интеллекта

Содержит информацию о каждой модели, разработанной для *СИИ* или используемой в ней (см. таблицу Е.13 и 6.5.4, 6.5.5).

Таблица Е.13 — Информация о модели

a)	Данные (например, обучающие, контрольные, тестовые) и алгоритмы, использованные для разработки моделей (см. 6.5.3)	Описание данных и алгоритмов, использованных для разработки модели
b)	Параметры и методы обучения модели	Описание параметров и методов обучения модели
c)	Образцы данных не используются повторно между обучающими, тестовыми и контрольными наборами данных	Описание шагов, предпринятых для обеспечения того, чтобы обучающие, контрольные и тестовые наборы данных были разделены
d)	Шаги, предпринятые для обучения, тестирования и контроля нескольких моделей	Описание шагов, предпринятых для разработки нескольких моделей
e)	Критерии и шаги, предпринятые для выбора подходящих признаков	Описание критериев и шагов, предпринятых для отбора признаков
f)	Критерии для выбора модели	Описание критериев, определенных для выбора модели
g)	Метрики, применяемые для оценки показателей деятельности моделей	Например, точность, рабочая характеристика приемника, матрица ошибок, среднеквадратичная ошибка или среднее расстояние до центра кластера
h)	Оценка того, насколько хорошо выбранные модели обобщают эксплуатационные данные	Описание оценок для определения того, обобщает ли модель эксплуатационные данные
i)	Оценка того, сохраняют или создают ли выбранные модели нежелательную предвзятость или другой вред	Описание оценок для определения того, сохраняют или создают ли модели нежелательную предвзятость или другой вред
j)	Оценка того, содержат или создают ли выходные данные модели ПДн	Описание оценок для определения того, содержат или создают ли модели ПДн
k)	Робастность модели	Описание метрик, принятых для оценки робастности моделей
l)	Обнаружение и коррекция дрейфа данных	Например, переобучение модели, если эксплуатационные данные начинают отличаться по характеристикам от исходных обучающих данных
m)	Критерии и процесс переобучения модели	Описание критериев и процессов переобучения моделей
n)	Воздействия непрерывного обучения (если внедрено)	Описание обоснованно предсказуемых воздействий в результате использования непрерывного обучения
o)	Вычислительные и экологические соображения	Описание того, как учитывались вычислительные и экологические аспекты
Юридические, организационные или технические требования		
1	Название и описание требования	Описание того, как было выполнено требование
2	Название и описание требования	Описание того, как было выполнено требование
…		…
Невыполненные юридические, организационные или технические требования		
1	Название и описание требования	Планы по выполнению требования
2	Название и описание требования	Планы по выполнению требования
…		…

Е.2.5 Среда развертывания

Е.2.5.1 Географический регион и языки

Содержит информацию о географическом, культурном или правовом контексте стран или регионов, где развернута *СИИ* (см. таблицу Е.14 и 6.6.1).

Таблица Е.14 — Среда развертывания

a1)	Текущий регион развертывания *СИИ*	Указание географических регионов, где в настоящее время развернута система
a2)	Регионы планируемого развертывания	Указание географических регионов, в которых планируется развертывание системы в будущем
b)	Особые требования законодательства в географическом регионе развертывания	Перечень требований законодательства, применимых в географическом регионе развертывания
c)	Культурные особенности географического региона развертывания	Описание соответствующих культурных особенностей географического региона развертывания
d)	Дискриминируемые группы, группы риска и маргинализированные группы	Описание дискриминируемых групп, групп риска или маргинализированных групп населения в географическом регионе развертывания соответственно оцениваемой *СИИ*
e)	Языки, на которых говорят в географическом регионе развертывания	Описание языка, на котором говорят в географическом регионе развертывания соответственно оцениваемой *СИИ*, в частности, в системах обработки естественного языка
f)	Поведенческие, социальные и физиологические черты человека, которые могут быть отражены в географическом положении	Описание поведенческих, социальных или физиологических черт как аспектов, релевантных для оцениваемой *СИИ*

Е.2.5.2 Сложность и ограничения среды развертывания

Включает информацию о технической среде развертывания *СИИ* (см. таблицу Е.15).

Таблица Е.15 — Сложность и ограничения среды развертывания

Каким образом выполнено или планируется к выполнению развертывание *СИИ*?	Информация о том, каким образом выполнено или планируется к выполнению развертывание оцениваемой *СИИ* согласно описанию в 6.6.2
Ограничения для развертывания	Описание ограничений для развертывания (если применимо)

Е.2.6 Соответствующие заинтересованные стороны

Определяется, кто является соответствующими заинтересованными сторонами, на которых потенциально может оказать воздействие *СИИ* и ее использование (см. таблицы Е.16 и Е.17, а также 6.7).

Таблица Е.16 — Внешние соответствующие заинтересованные стороны

Номер	Внешняя соответствующая заинтересованная сторона	Роль по отношению к *СИИ*
1	Отдельные лица, группы лиц, уязвимые группы, социальные группы, субъекты данных и т. д.	Описание влияния, которое *СИИ* может оказать на установленную заинтересованную сторону
2	Отдельные лица, группы лиц, уязвимые группы, социальные группы, субъекты данных и т. д.	Описание влияния, которое *СИИ* может оказать на идентифицированную заинтересованную сторону
3	…	…

Таблица Е.17 — Внутренние соответствующие заинтересованные стороны

Номер	Внутренняя соответствующая заинтересованная сторона или команда	Роль по отношению к *СИИ* (например, разработчик, владелец системы или ответственный за надзор)
1	Команда, отдел и т. д.	Описание роли по отношению к *СИИ* и какому потенциальному влиянию подвергается
2	Команда, отдел и т. д.	Описание роли по отношению к *СИИ* и какому потенциальному влиянию подвергается
3	…	…

Е.2.7 Фактические и обоснованно предсказуемые польза и вред

Определяет фактические и обоснованно предсказуемые пользу и вред для каждой идентифицированной заинтересованной стороны, указанной в Е.2.6, с использованием точек зрения, описанных в 6.8.2 (см. таблицу Е.18).

Таблица Е.18 — Польза и вред

Наименование показателя	Польза	Вред
Подотчетность	Описание обоснованно предсказуемой пользы в отношении подотчетности	Описание обоснованно предсказуемого вреда в отношении подотчетности
Прозрачность	Описание обоснованно предсказуемой пользы в отношении прозрачности	Описание обоснованно предсказуемого вреда в отношении прозрачности
Справедливость и дискриминация	Описание обоснованно предсказуемой пользы в отношении справедливости и дискриминации	Описание обоснованно предсказуемого вреда в отношении справедливости и дискриминации
Конфиденциальность	Описание обоснованно предсказуемой пользы в отношении конфиденциальности	Описание обоснованно предсказуемого вреда в отношении конфиденциальности
Надежность	Описание обоснованно предсказуемой пользы в отношении надежности	Описание обоснованно предсказуемой пользы в отношении надежности
Безопасность	Описание обоснованно предсказуемой пользы в отношении безопасности	Описание обоснованно предсказуемой пользы в отношении безопасности
Объяснимость	Описание обоснованно предсказуемой пользы в отношении объяснимости	Описание обоснованно предсказуемой пользы в отношении объяснимости
Воздействие на окружающую среду	Описание обоснованно предсказуемой пользы в отношении воздействия на окружающую среду	Описание обоснованно предсказуемой пользы в отношении воздействия на окружающую среду

Е.2.8 Сбои системы искусственного интеллекта и обоснованно предсказуемое нецелевое использование

Е.2.8.1 Воздействие сбоя системы искусственного интеллекта

Для каждой из заинтересованных сторон, указанных в Е.2.6, приводится информация о воздействии сбоя *СИИ* (см. таблицу Е.19 и 6.8.3.2).

Таблица Е.19 — Воздействия сбоев в *СИИ*

Номер	Описание сбоя	Воздействия, вызванные потенциальным сбоем (указывают, польза или вред)
1	Описание потенциальных сбоев, которые могут оказать воздействие на соответствующие заинтересованные стороны	Описание воздействий, вызванных идентифицированными потенциальными сбоями
2	Описание потенциальных сбоев, которые могут оказать воздействие на соответствующие заинтересованные стороны	Описание воздействий, вызванных идентифицированными потенциальными сбоями
3	…	…

Е.2.8.2 Воздействие обоснованно предсказуемого нецелевого использования системы искусственного интеллекта

Для каждой из заинтересованных сторон, указанных в Е.2.6, приведена информация о воздействии обоснованно предсказуемого нецелевого использования оцениваемой *СИИ* (см. таблицу Е.20 и 6.8.3.3).

Таблица Е.20 — Воздействие обоснованно предсказуемого нецелевого использования *СИИ*

Номер	Описание обоснованно предсказуемого нецелевого использования	Воздействия в результате обоснованно предсказуемого нецелевого использования
1	Описание обоснованно предсказуемого нецелевого использования, которое может оказать воздействие на соответствующие заинтересованные стороны	Описание воздействий, вызванных выявленным нецелевым использованием (указывают, польза или вред)
2	Описание обоснованно предсказуемого нецелевого использования, которое может оказать воздействие на соответствующие заинтересованные стороны	Описание воздействий, вызванных выявленным нецелевым использованием (указывают, польза или вред)
3	…	…

Приложение ДА
(справочное)

Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте

Таблица ДА.1

Обозначение ссылочного национального стандарта	Степень соответствия	Обозначение и наименование ссылочного международного стандарта
ГОСТ Р 71476—2024 (ИСО/МЭК 22989:2022)	MOD	ISO/IEC 22989:2022 «Информационные технологии. Искусственный интеллект. Концепции и терминология искусственного интеллекта»
Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандарта: - MOD — модифицированный стандарт.		

Библиография

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

[2] ИСО/МЭК 23894:2023 Информационная технология. Искусственный интеллект. Руководство по менеджменту риска (Information technology — Artificial intelligence — Guidance on risk management)

[3] *ИСО/МЭК 23053:2022* *Экосистема разработки систем искусственного интеллекта (ИИ) с использованием машинного обучения (МО) [Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)]*

[4] ИСО 10377:2013 Безопасность потребительских товаров. Руководящие указания для поставщиков (Consumer product safety — Guidelines for suppliers)

[5] *Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных»*

[6] *Федеральный закон от 26 июля 2017 г. № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации»*

[7] ИСО/МЭК 23751:2022 Информационная технология. Облачные вычисления и распределенные платформы. Рамочное соглашение об обмене данными (DSA) [Information technology — Cloud computing and distributed platforms — Data sharing agreement (DSA) framework]

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

[9] ISO/IEC TS 12791:2024 Информационная технология. Искусственный интеллект. Устранение нежелательных отклонений в задачах машинного обучения классификации и регрессии [Information technology — Artificial intelligence (AI) — Treatment of unwanted bias in classification and regression machine learning tasks]

[10] ИСО/МЭК 29134:2023 Информационная технология. Методы и средства обеспечения безопасности. Руководящие указания по оценке воздействия на конфиденциальность (Information technology — Security techniques — Guidelines for privacy impact assessment)

[11] ИСО/МЭК 27701 Информационная безопасность, кибер-безопасность и защита частной жизни. Системы управления конфиденциальной информацией. Требования и руководство (Information security, cybersecurity and privacy protection — Privacy information management systems — Requirements and guidance)

[12] ISO/IEC TR 20226 Информационные технологии. Искусственный интеллект. Аспекты экологической устойчивости систем искусственного интеллекта (Information technology — Artificial intelligence — Environmental sustainability aspects of AI systems)

УДК 004.8:006.354 ОКС 35.020

Ключевые слова: искусственный интеллект, система искусственного интеллекта, оценка воздействия системы искусственного интеллекта

Редактор *Е.В. Якубова*
Технический редактор *И.Е. Черепкова*
Корректор *Е.Д. Дульнева*
Компьютерная верстка *М.В. Малеевой*

Сдано в набор 11.02.2026. Подписано в печать 27.02.2026. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 5,12. Уч.-изд. л. 4,35.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru