ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

ПРЕДВАРИТЕЛЬНЫЙ НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПНСТ 963— 2024 (ИСО/МЭК 5339:2024)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Руководство для приложений на основе искусственного интеллекта

(ISO/IEC 5339:2024, Information technology — Artificial intelligence — Guidance for AI applications, MOD)

Издание официальное

Москва
Российский институт стандартизации
2024

Предисловие

1 ПОДГОТОВЛЕН Федеральным государственным автономным образовательным учреждением высшего образования «Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ) 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 25 октября 2024 г. № 70-пнст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 5339:2024 «Информационные технологии. Искусственный интеллект. Руководство по применению искусственного интеллекта» (ISO/IEC 5339:2024 «Information technology — Artificial intelligence — Guidance for AI applications», MOD) путем изменения отдельных фраз (слов, значений показателей, ссылок), которые выделены в тексте курсивом.

Внесение указанных технических отклонений направлено на учет особенностей российской национальной стандартизации.

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

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

Правила применения настоящего стандарта и проведения его мониторинга установлены в ГОСТ Р 1.16—2011 (разделы 5 и 6).

Федеральное агентство по техническому регулированию и метрологии собирает сведения о практическом применении настоящего стандарта. Данные сведения, а также замечания и предложения по содержанию стандарта можно направить не позднее чем за 9 мес до истечения срока его действия разработчику настоящего стандарта по адресу: info@tc164.ru и/или в Федеральное агентство по техническому регулированию и метрологии по адресу: 123112 Москва, Пресненская набережная, д. 10, стр. 2.

В случае отмены настоящего стандарта соответствующая информация будет опубликована в ежемесячном информационном указателе «Национальные стандарты» и также будет размещена на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (www.rst.gov.ru)

Содержание

Введение

Системы искусственного интеллекта (ИИ) обладают потенциалом для создания постепенных изменений и достижения новых уровней производительности и возможностей в таких сферах, как сельское хозяйство, транспорт, финансовые технологии, образование, энергетика, здравоохранение и промышленное производство. Однако потенциальные риски, связанные с низкой степенью доверия, могут повлиять на применение и распространение ИИ. Приложения на основе ИИ могут оказывать влияние на множество заинтересованных сторон, включая отдельных лиц, организации и общество в целом. Воздействие от использования ИИ может изменяться с течением времени, в некоторых случаях это связано с особенностями используемых данных или с изменениями нормативной правовой среды. До заинтересованных сторон следует донести информацию об особенностях их роли и степени их ответственности. Если детальное описание проблем в соответствующих стандартах создается в интересах технических специалистов, в настоящем стандарте на макроуровне описан контекст жизненного цикла приложения на основе ИИ с целью способствовать взаимопониманию разных сторон, их заинтересованности и одобрению технологии ИИ сообществом.

Настоящий стандарт содержит рекомендации в отношении приложений на основе ИИ, отражающие различные точки зрения всех заинтересованных сторон на макроуровне. Стандарт охватывает точки зрения на «разработку», «использование» и «воздействие» ИИ, в нем рассмотрены функциональные, а также нефункциональные характеристики, такие как степень доверия к ИИ и управление рисками. Руководство может быть использовано разработчиками стандартов, создателями приложений и другими заинтересованными сторонами для ответа на вопрос о характеристиках приложений на основе ИИ и для решения других заслуживающих внимания вопросов. Заинтересованные стороны представлены на разных стадиях жизненного цикла системы ИИ, выделены их роли и сферы ответственности, продемонстрированы процессы, которые должны произойти на последующих этапах, для того, чтобы участие заинтересованных сторон соответствовало их роли. Заинтересованные стороны могут иметь разные уровни профессиональных знаний и опыта. Поскольку приложения на основе ИИ могут отличаться от программных продуктов, не использующих ИИ, за счет возможностей постоянного развития ИИ и необходимости учета степени доверия к технологии, всем заинтересованным сторонам следует быть информированными о специфических характеристиках ИИ.

Настоящий стандарт включает:

- общие положения, цели и задачи (раздел 4);
- методику определения заинтересованных сторон, контекста использования, функциональных и нефункциональных характеристик приложений на основе ИИ (раздел 5);
- принципиальные подходы для ответа на вопрос о характеристиках и заслуживающих рассмотрения аспектах в отношении приложений на основе ИИ (раздел 6);
- рекомендации для приложений на основе ИИ с точек зрения «разработки», «использования» и «воздействия» (раздел 7).

ПРЕДВАРИТЕЛЬНЫЙ НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Руководство для приложений на основе искусственного интеллекта

Artificial intelligence.
Guidance for artificial intelligence applications

Срок действия — с 2025—01—01
до 2027—01—01

1 Область применения

В настоящем стандарте содержатся рекомендации в отношении условий использования, возможностей и процессов при разработке и применении приложений на основе искусственного интеллекта (ИИ). Стандарт представляет собой взгляд на макроуровне на контекст применения ИИ, на заинтересованные стороны, их роли и взаимосвязи с жизненным циклом системы, а также на типичные характеристики и особенности применения ИИ.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 70462.1 (ISO/IEC TR 24029-1—2021) Информационные технологии. Интеллект искусственный. Оценка робастности нейронных сетей. Часть 1. Обзор

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

ГОСТ Р 71484.2 (ИСО/МЭК 5259-2:2023) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Меры качества данных

ГОСТ Р 71484.3 (ИСО/МЭК 5259-3:2023) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 3. Требования и руководство по управлению качеством данных

ГОСТ Р 71484.4 (ИСО/МЭК 5259-4:2023) Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 4. Структура процесса управления качеством данных

ГОСТ Р 71539 (ИСО/МЭК 5338:2023) Искусственный интеллект. Процессы жизненного цикла систем искусственного интеллекта

ГОСТ Р ИСО 26000 Руководство по социальной ответственности

ПНСТ 837—2023 Искусственный интеллект. Управляемость автоматизированных систем искусственного интеллекта

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 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

Издание официальное

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*, а также следующие термины с соответствующими определениями:

3.1

применение искусственного интеллекта; применение ИИ (AI application): Использование технологии ИИ с определенными функциональными характеристиками для достижения ожидаемого результата в условиях, заданных заинтересованной стороной.

Примечание — В зависимости от контекста в тексте стандарта используется также термин «приложение на основе ИИ».

3.2

служба облачных вычислений (cloud service): Одна или более возможностей, предоставляемых через облачные вычисления (см. 3.2.5), вызываемая посредством определенного интерфейса.

[ГОСТ ISO/IEC 17788—2016, статья 3.2.8]

3.3

частное облако (private cloud): Модель развертывания облачных вычислений (см. 3.2.7), в которой службы облачных вычислений (см. 3.2.8) используются исключительно единственным потребителем службы облачных вычислений (см. 3.2.11), и ресурсами управляет тот же потребитель службы облачных вычислений (см. 3.2.11).

[ГОСТ ISO/IEC 17788—2016, статья 3.2.32]

3.4

потребитель службы облачных вычислений (cloud service customer): Сторона (см. 3.1.6), которая находится в деловых отношениях в целях использования служб облачных вычислений (см. 3.2.8).

Примечание — Деловые отношения не обязательно подразумевают финансовые соглашения.

[ГОСТ ISO/IEC 17788—2016, статья 3.2.11]

3.5

модель развертывания облачных вычислений (cloud deployment model): Способ организации облачных вычислений (см. 3.2.5), основанный на управлении и совместном использовании физических или виртуальных ресурсов.

Примечание — Модели развертывания облачных вычислений включают в себя общественное облако (см. 3.2.19), гибридное облако (см. 3.2.23), частное облако (см. 3.2.32) и публичное облако (см. 3.2.33)

[ГОСТ ISO/IEC 17788—2016, статья 3.2.7]

3.6

облачные вычисления (cloud computing): Парадигма для предоставления возможности сетевого доступа к масштабируемому и эластичному пулу общих физических или виртуальных ресурсов с предоставлением самообслуживания и администрированием по требованию.

Примечание — Примеры ресурсов включают серверы, операционные системы, сети, программное обеспечение, приложения и оборудование для хранения данных.

[ГОСТ ISO/IEC 17788—2016, статья 3.2.5]

4 *Общие положения. Цели и задачи*

В настоящем стандарте рассматриваются характеристики и другие особенности и факторы, требующие анализа применительно к приложениям на основе ИИ. Это является основой для взаимопонимания между заинтересованными сторонами, способствующей информационному обмену, заинтересованности в применении, в целом принятию и одобрению технологий ИИ.

Данный стандарт определяет:

- контекст приложения на основе ИИ с учетом аспектов: «кто» (заинтересованные стороны), «что», «когда», «где», «для чего», «как» (для разных стадий жизненного цикла ИИ);

- *заинтересованные стороны: разработчик, провайдер, производитель ИИ, заказчик (клиент), партнер, пользователи, поставщик данных, сообщество и соответствующие государственные органы;*

- типовые функциональные и нефункциональные характеристики и заслуживающие внимания аспекты для анализа приложения на основе ИИ.

5 Контекст применения системы ИИ

5.1 Определение подхода к контексту применения системы ИИ

В настоящем разделе изложен подход к определению контекста применения системы ИИ. Стадии жизненного цикла системы ИИ рассмотрены в соответствии с *ГОСТ Р 71476—2024* (раздел 6) и *ГОСТ Р 71539*. На каждой стадии заинтересованные стороны, процессы и взаимосвязи определены через следующие аспекты:

- «кто» — заинтересованные стороны (например, организации, физические лица или группы лиц), связанные с данным контекстом, чьи интересы и ценности могут быть учтены и чьи проблемы могут быть решены;

- «что» — деятельность, связанная с условиями применения ИИ, такая как:
  - возможности системы ИИ и ее приложения;
  - типы принимаемых решений, поддерживаемых данным приложением;

- «как» — конкретные методы в конкретном контексте, такие как:
  - степень участия человека в процессе принятия решений (например, автономный процесс или полуавтономный);
  - система ИИ в роли дополнительного инструмента (например, поддержка при принятии решения, взаимодействие человека и системы ИИ);
  - алгоритмические процессы;
  - источники данных, сбор и предоставление данных;
  - ввод в действие системы в виде продукта или услуги;

- «когда» — параметр времени, связанный с данным контекстом, т. е. с некоторым процессом на определенной стадии жизненного цикла системы ИИ, или с моментом времени активации процесса, например с частотой применения. Это зависит от контекста, заданного аспектом «что»;

- «где» — локация, связанная с контекстом, то есть, где именно приложение на основе ИИ используется: внутри организации (например, в рабочих процессах) или вне организации (например, в работе с клиентами); режим ввода в действие приложения (например, локально, как служба облачных вычислений, или через третьих лиц);

- «для чего» — внешние причины и мотивация, связанные с данным контекстом, то есть элемент привлекательности для аспекта «кто», то есть для клиентов, пользователей и общества в целом, также целесообразность приложения, его цели, преимущества, ограничения и воздействие на экономическую, социальную и другие сферы.

5.2 Контекст применения системы ИИ

На рисунке 1 показан типичный контекст применения ИИ с заинтересованными сторонами, процессами и взаимосвязями с различными стадиями жизненного цикла системы ИИ.

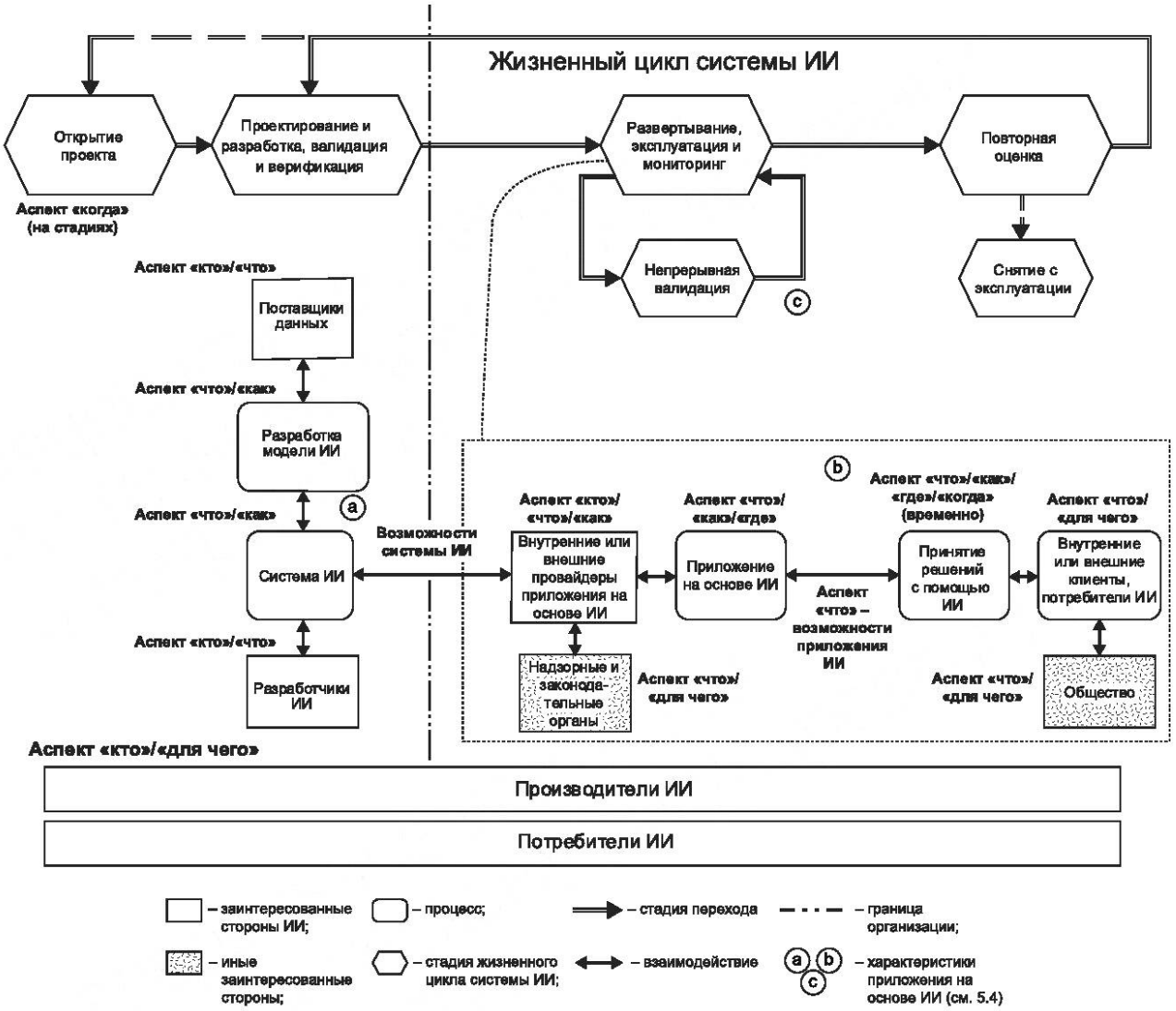

Рисунок 1 — Типичный контекст применения ИИ

К другим заинтересованным сторонам относятся представители сообщества, которые не участвуют в разработке или использовании приложения ИИ, но подвержены воздействию от его применения, а также регуляторы и законодательные органы, которые оказывают влияние на его внедрение.

Взаимоотношения между заинтересованными сторонами включают в себя информационное взаимодействие и обмен мнениями. Организационные границы используются для определения того, что находится внутри ведения организации производителя, а что — за ее пределами (например, до и после развертывания). В некоторых случаях поставщик приложения ИИ может быть частью организации-производителя, но при этом играть внешнюю роль. Три характеристики приложений ИИ (см. 5.4) также отражены на рисунке 1.

5.3 Заинтересованные стороны и процессы

5.3.1 Общие положения

На рисунке 1 показаны взаимосвязи между заинтересованными сторонами (аспект «кто»), их ролями (аспекты «что», «где», «когда») и процессами (аспект «что»), которые используются (аспект «как»).

На рисунке 1 также показано, что производитель, потребитель, регулирующие органы и сообщество (аспект «кто») также руководствуются ценностными соображениями (аспект «для чего») в данном контексте.

5.3.2 Стороны, заинтересованные в использовании ИИ

5.3.2.1 Общие положения

Стороны, заинтересованные в использовании ИИ, приведенные в настоящей статье, играют одну (основную или второстепенную) роль или более, на разных стадиях жизненного цикла системы ИИ. Название заинтересованной стороны связано с ее основной или второстепенной ролью в соответствии с *ГОСТ Р 71476—2024* (статья 5.19).

5.3.2.2 Производитель ИИ

Производитель ИИ (аспект «кто») — это организация или компания, которая проектирует, создает, тестирует и вводит в действие продукты или сервисы, использующие одну систему ИИ или более. Производитель ИИ принимает на себя эти роли в соответствии с целями организации (аспект «для чего», например, для извлечения прибыли или создания дополнительной стоимости для клиентов). Эти роли охватывают все стадии жизненного цикла системы ИИ (аспект «когда»), включая принятие решений о проектировании, завершении проекта и выводе системы из эксплуатации.

5.3.2.3 Разработчик ИИ

Разработчик ИИ — это организация или компания, которая занимается разработкой продуктов и сервисов на основе ИИ для производителя. Ее роли включают проектирование модели и системы ИИ, разработку, внедрение, верификацию и валидацию (аспект «что») на стадии, предшествующей вводу в действие в жизненном цикле системы (аспект «когда»). Отдельно взятый разработчик может работать в штате организации-производителя, по контракту или на правах партнера.

5.3.2.4 Заказчик (клиент) ИИ

Клиент, использующий ИИ — это организация или компания, которая использует продукт или сервис либо непосредственно в своей работе, либо в роли провайдера для пользователей. Между провайдером приложения ИИ (см. 5.3.2.6) и клиентом, его использующим, существуют деловые связи, например установленные деловые отношения, покупка продукта или подписка на услуги. Роль клиентов соотносится с жизненным циклом системы ИИ (аспект «когда»), поскольку клиенты создают спрос, реализуют стоимость и поддерживают жизнеспособность конкретного продукта ИИ (аспект «для чего»). С клиентами часто консультируются на предварительной стадии проектирования для установления требований к продукту, они также принимают участие на стадии верификации и валидации, внедрении системы, в эксплуатации и мониторинге системы и на стадии вывода системы из эксплуатации.

Клиент, использующий ИИ, или пользователь ИИ (см. 5.3.2.5) может быть в штате организации-провайдера приложения на основе ИИ (например, подразделение организации, ответственное за бизнес-функцию) или может иметь с провайдером опосредованную связь — внешний пользователь, например, если провайдер предоставляет услуги третьей стороне (аспект «где»).

5.3.2.5 Пользователь ИИ

Пользователь ИИ — это организация или лицо, которые используют продукты или услуги на основе ИИ. Пользователь ИИ может быть индивидуальным представителем некоторого сообщества (аспект «кто») или представителем организации или компании-клиента. Клиент может быть одновременно и пользователем. Пользователь ИИ не обязательно должен быть клиентом, то есть ему не обязательно иметь деловые отношения с провайдером ИИ (см. 5.3.2.6). Роль пользователя ИИ обычно соотносится со стадиями эксплуатации и мониторинга в жизненном цикле системы ИИ (аспект «когда»), во время которых возникает отдача за счет применения продукта или сервиса ИИ (аспект «для чего»).

5.3.2.6 Провайдер приложения на основе ИИ

В целом провайдер приложения на основе ИИ — это организация, которая предоставляет пользователям продукты или услуги, используя одну систему ИИ или более. В контексте применения ИИ провайдер приложения на основе ИИ (аспект «кто») — это организация, которая предоставляет возможности системы ИИ (такие, как построение логического вывода и принятие решений) в форме приложения на основе ИИ (аспект «что») в качестве продукта или услуги (аспект «как») для клиентов внутри организации или вне ее в соответствии с *ГОСТ Р 71476*.

Примечание — Провайдер приложения на основе ИИ в настоящем стандарте аналогичен провайдеру продукта или услуг ИИ по *ГОСТ Р 71476*.

Провайдер приложения на основе ИИ может быть внутренним (в штате организации-производителя ИИ) или внешним (третья сторона в качестве провайдера продукта или сервиса).

Роль провайдера ИИ обычно соотносится со стадиями ввода в действие в жизненном цикле системы ИИ (аспект «когда»). Провайдер может принимать участие на более ранних стадиях, предостав-

ляя данные о потенциальных сферах применения, локациях, клиентах и пользователях, типах решений и специфике среды для внедрения системы.

5.3.2.7 Партнер в сфере ИИ

Партнер в сфере ИИ — это организация или компания, которая предоставляет услуги производителю ИИ или провайдеру приложения на основе ИИ в рамках деловых отношений.

5.3.2.8 Поставщик данных

Поставщик данных (аспект «кто») — это организация или компания, которая обеспечивает поставку данных, используемых для продуктов или сервисов ИИ. Поставщик данных осуществляет сбор данных или их подготовку (аспект «что») либо выполняет и то, и другое для использования моделью ИИ производителя. Поставщик данных может быть партнером производителя ИИ.

Роль поставщика данных соотносится со стадиями до ввода в действие системы (аспект «когда»). В определенных обстоятельствах, например, если в системе ИИ используются модели машинного обучения, поставщик данных может быть задействован и после внедрения системы для сбора и подготовки данных для непрерывной валидации (аспект «когда»).

5.3.3 Иные заинтересованные стороны

5.3.3.1 Общие положения

Иные заинтересованные стороны являются представителями сообщества, не занятыми разработкой или использованием приложения на основе ИИ, но испытывающие эффект от его применения, например, потребители, надзорные и законодательные органы, чьи полномочия могут влиять на условия использования ИИ, также относятся к этой категории.

5.3.3.2 Общественность

Использование технологии ИИ может влиять не только на отдельного клиента или пользователя, но и на других членов сообщества (аспект «кто»), например, потребителей, членов семьи, соседей, коллег по работе, знакомых, родственников.

5.3.3.3 Надзорные и законодательные органы

Надзорный орган (аспект «кто») — орган власти на определенной территории, где приложение с использованием ИИ разворачивается и используется. Этот орган имеет юрисдикцию над использованием технологии ИИ на основе существующих требований закона. Несмотря на то, что соответствие закону определяется надзорными органами на стадиях ввода в действие системы, ее эксплуатации и при мониторинге, провайдеру ИИ и другим сторонам, чьи интересы затрагиваются на начальных стадиях жизненного цикла системы, следует выявить релевантные риски с точки зрения надзорных органов и найти решения, чтобы избежать препятствий на пути к поставленной цели.

Приложения на основе ИИ могут вводиться в действие в юрисдикциях, имеющих разные правила, касающиеся сбора и использования данных и работы с ними.

Законодательный орган (аспект «кто») — орган власти на определенной территории, где приложение с использованием ИИ внедряется и функционирует, этот орган устанавливает правовые нормы, управляющие использованием технологии ИИ.

5.3.4 Процессы

5.3.4.1 Общие положения

Процесс — это функция или деятельность, преобразующая определенные входные данные в данные заданного типа на выходе. С точки зрения приложения на основе ИИ, процессы, описанные в данной статье, относятся к входным данным (аспект «что»), которые преобразуются (аспект «как») системой ИИ в ее возможности (аспект «что»). Эти возможности внедряются (аспекты «как», «где») в приложении на основе ИИ (аспект «что») в качестве дополнительного инструмента принятия решений пользователем (аспекты «что», «как», «когда»).

5.3.4.2 Система ИИ

Система ИИ — это специальная система, спроектированная, созданная, верифицированная и валидированая производителем ИИ для выполнения определенных функций, таких, как построение логического вывода и принятие решений в соответствии с *ГОСТ Р 71476*. Эти функции определяют «что может сделать система ИИ», то есть возможности системы ИИ. Каким образом такие возможности достигнуты, зависит от конфигурации и структуры модели ИИ (см. 5.3.4.3) и вида деятельности, для которой она предназначена, например машинное зрение, распознавание изображений, обработка текстов на естественном языке, машинный перевод, синтез речи, интеллектуальный анализ данных и планирование.

5.3.4.3 Модель ИИ и ее разработка

Модель ИИ — это представление в математической форме некоторого процесса (аспект «что»), оно формирует основу системы ИИ (см. 5.3.4.2). Модель ИИ может создаваться с помощью разных технологий, таких как нейронные сети, деревья решений, байесовские сети, логические высказывания и онтологии. Эти модели используются с целью прогнозирования или вычисления решений для выполнения функций системы ИИ в соответствии с *ГОСТ Р 71476—2024* (пункт 8.3).

Данные, требуемые для разработки модели, могут быть получены с помощью машинного обучения путем обработки подготовленных данных с помощью алгоритма (аспект «как»). Требуемые данные получают из источников, релевантных для конкретной сферы деятельности и среды принятия решений приложения, использующего ИИ (см. 5.3.4.4). В качестве альтернативного варианта данные могут извлекаться также из накопленных человеком декларативных или процедурных знаний, а человеческий опыт может использоваться в логическом программировании и экспертных системах, основанных на правилах, для получения логического вывода (аспект «как») (см. *[1]*).

5.3.4.4 Приложение на основе ИИ

Возможности системы ИИ применяются в конкретной среде принятия решений, в конкретной сфере деятельности, включая сельское хозяйство, транспортную систему, финансовые технологии, образование, энергетику, здравоохранение, промышленное производство и многие другие отрасли. Такое приложение может включать иные, не связанные с использованием ИИ, характеристики и особенности, разработанные для конкретного клиента, чтобы удовлетворить требования конкретной системы и потребности клиентов или пользователей. Провайдер объединяет возможности системы ИИ в приложение, которое может быть введено в действие в конкретной среде, и это приложение, в свою очередь, демонстрирует в этой среде свои уникальные возможности. Провайдер приложения на основе ИИ может внедрить его как продукт или как сервис. Сведения об уровнях автоматизации приложения на основе ИИ приведены в 5.3.4.6 и *ГОСТ Р 71476—2024* (подраздел 5.13).

Одним из примеров является система ИИ, которая обладает возможностями распознавания естественного языка, которые сочетаются с возможностями построения диалога и функцией анализа эмоциональной окраски высказываний, в результате получается чат-бот, развернутый в виде службы облачных вычислений для взаимодействия с пользователями онлайн (см. пример использования в А.3). Другой пример — система ИИ, имеющая возможности распознавания изображений в сочетании с глубоким обучением, используемая в медицинской диагностике в виде приложения, обнаруживающего аномалии в изображениях биологических объектов. Для такого приложения при обучении и оценке результатов используется визуальная информация, которая формируется в соответствии с характеристиками системы и интерфейса для специалистов по клинической диагностике, которые распознают изображения.

5.3.4.5 Услуги по предоставлению приложений на основе ИИ

Услуги на основе приложений в области ИИ — это деятельность, выполняемая для заказчика или пользователя ИИ (аспект «как»), которая основана на возможностях приложения ИИ. Внедрение услуг может быть локальным (например, А.2) или в облачном сервисе (аспект «где») (например, А.3).

5.3.4.6 Принятие решений с использованием ИИ

В типовой ситуации принятия решения человек, принимающий решение, оперирует следующими факторами:

- совокупностью неопределенных событий, каждое из которых может произойти с некоторой вероятностью;
- совокупностью действий, которые могут быть предприняты в случае, если события произойдут;
- совокупностью результатов предпринятых действий и определенных событий, которые происходят в реальности.

Для снижения неопределенности в прогнозировании событий человек, принимающий решения, может попытаться получить более точную оценку вероятности того, что события произойдут. Это можно сделать путем сбора релевантной информации о соответствующей среде и обработке этих данных в контексте конкретного решения для получения прогнозов. На основе этих прогнозов и критериев для принятия решения, например максимизации среднего ожидаемого значения решения, знания человека, принимающего решение, могут использоваться, чтобы определить, какое действие предпринять, исходя из среднего ожидаемого значения результата, и затем выбрать, как действовать наилучшим образом. В некоторых случаях информация из внешних источников также принимается во внимание для принятия окончательного решения.

Приложение на основе ИИ является инструментом, который использует человек, принимающий решение, для выполнения стоящей перед ним задачи. Как показано на рисунке 1, сбор и подготовка данных, которые затем обрабатываются конкретной моделью ИИ (см. 5.3.4.3), осуществляется для конкретной системы ИИ (см. 5.3.4.2). Эта модель делает прогнозы, которые могут помочь в процессе принятия решения и выбора плана действий.

Приложение на основе ИИ создается таким образом, чтобы функционировать в некоторой степени автоматически. Первый вариант — решение принимается и действие выполняется самим приложением без вмешательства человека, второй — использование рекомендации, выданной приложением, чтобы дополнить знания и опыт человека, который принимает конечное решение и осуществляет выбранное действие. Решение может быть принято, когда это надлежит по плану, по сигналу от сенсора или при наступлении определенного события, когда такое решение потребуется (аспект «когда»).

5.4 Функциональные характеристики приложения на основе ИИ

Приложение, использующее ИИ, отличается от приложения, которое не использует ИИ, одной или несколькими функциональными характеристиками, приведенными ниже:

- приложение на основе ИИ создают на базе возможностей системы ИИ, которая задействует модель для получения и обработки данных без вмешательства человека с помощью алгоритма и программ. Модель создают на основе машинного обучения с учителем, без учителя, с частичным привлечением учителя или на основе программируемых правил. Сбор информации для построения модели может также включать такие процессы, которые связаны с применением полученной информации;
- приложение на основе ИИ применяет оптимизацию или логические выводы, полученные моделью, для получения вовремя улучшенных решений, прогнозов и рекомендаций, чтобы достичь поставленных конкретных целей. Дополнительные возможности, характеристики и уникальные особенности обычно добавляются в соответствии с требованиями конкретно взятой сферы применения или среды принятия решений или по запросу клиентов или пользователей. Форма применяемой оптимизации результата логического вывода зависит от конкретной создаваемой модели. Выходные результаты используют как дополнение информации пользователя при принятии решений, прогнозировании и создании рекомендаций. Приложение может реагировать и выдавать ответную информацию в изменяющейся среде, включая функционирование в реальном времени;
- приложение на основе ИИ в некоторых случаях обновляется. Модель, систему и приложение совершенствуют путем оценки результатов взаимодействия с окружающей средой. Результаты взаимодействия с пользователями оценивают на основе метрики производительности модели и затем данные используют для непрерывного обучения и совершенствования.

Связь между тремя перечисленными характеристиками и заинтересованными сторонами в контексте применения ИИ показана на рисунке 1.

5.5 Нефункциональные характеристики приложения на основе ИИ

5.5.1 Общие положения

Функциональные характеристики приложения на основе ИИ описаны в 5.4. Нефункциональные характеристики такого приложения также следует учитывать заинтересованным сторонам при принятии решений, касающихся приложения на основе ИИ. Настоящий подраздел посвящен нефункциональным требованиям, специфичным для приложений на основе ИИ.

5.5.2 Доверие к системе ИИ

5.5.2.1 Общие положения

Доверие к системе ИИ — нефункциональная и существенная характеристика системы ИИ. Это означает, что система соответствует ожиданиям заинтересованных сторон и степень соответствия может быть верифицирована в соответствии с *ГОСТ Р 71476—2024* (подраздел 5.15), при этом качество системы определяется как гарантоспособное и надежное (см. [2]). Доверие к приложению на основе ИИ включает в себя доверие к системе ИИ и дополнительным возможностям, особенностям и характеристикам, заданным клиентом. Разные взгляды на доверие к системе ИИ (ее создание, использование и воздействие) заинтересованных сторон описаны в разделе 7.

Подробная информация о доверии к системам ИИ приведена в *[2]* и *[3]*.

5.5.2.2 Робастность ИИ

Робастность ИИ — это способность системы ИИ при любых обстоятельствах поддерживать такой уровень производительности, который был задан ее разработчиками и который востребован ее

клиентами и пользователями. Дополнительную информацию о робастности нейронных сетей см. в *ГОСТ Р 71476—2024* (пункт 5.15.2), *ГОСТ Р 70462.1*.

5.5.2.3 Надежность ИИ

Надежность ИИ — это способность системы ИИ или любого ее субкомпонента выполнять требуемые функции в заданных условиях в течение определенного периода времени [см. *ГОСТ Р 71476—2024* (пункт 5.15.3)].

5.5.2.4 Восстанавливаемость системы ИИ

Восстанавливаемость — это способность системы ИИ быстро восстанавливать работоспособное состояние после отказа или нарушения в работе. Некоторые устойчивые системы могут после сбоя продолжать функционировать, хотя их возможности снижаются [см. *ГОСТ Р 71476—2024* (пункт 5.15.3)].

5.5.2.5 Управляемость

Управляемость — это свойство системы ИИ, при котором внешний агент может вмешиваться в ее работу. Управляемость может быть достигнута путем предоставления надежных механизмов, с помощью которых агент может взять на себя контроль над системой ИИ [см. *ГОСТ Р 71476—2024* (пункт 5.15.5)].

Управляемость для пользователя (user controllability) — степень, в которой пользователь может надлежащим образом своевременно вмешиваться в работу системы ИИ [см. *ПНСТ 837—2023* (пункт 3.21)].

5.5.2.6 Объяснимость

Объяснимость — это свойство системы ИИ, которое означает, что важные факторы, влияющие на решение, могут быть выражены понятным для человека способом [см. *ГОСТ Р 71476—2024* (пункт 5.15.6)].

5.5.2.7 Предсказуемость

Предсказуемость — это характеристика системы ИИ, которая позволяет заинтересованным сторонам делать надежные предположения, касающиеся поведения системы и данных, получаемых на выходе, в соответствии с *ГОСТ Р 71476—2024* (пункт 5.15.7). В *[2]* эта проблема рассмотрена с точки зрения непредсказуемости.

5.5.2.8 Прозрачность

Прозрачность позволяет заинтересованным сторонам получать информацию о том, для чего предназначена система ИИ, каким образом она разрабатывалась и вводилась в действие [см. *ГОСТ Р 71476—2024* (пункт 5.15.8)]. Сюда включена информация о целях, ограничениях, определениях, допущениях, алгоритмах, источниках и сборе данных, безопасности, защите частной информации и конфиденциальности и уровне автоматизации. Прослеживаемость (traceability) как элемент прозрачности в качестве потенциального источника этических проблем рассмотрена в *[4]*.

5.5.2.9 Верификация и валидация

Верификация — это подтверждение, что система ИИ была создана корректно и соответствует заданным требованиям. Валидация — это подтверждение с помощью объективных фактов, что требования, соответствующие заданному применению приложения на основе ИИ, были выполнены [см. *ГОСТ Р 71476—2024* (подраздел 5.16)]. В *[3]* описаны методы программной инженерии, применимые для валидации и верификации систем ИИ.

5.5.2.10 Предвзятость системы ИИ и справедливость

Поведение системы ИИ, созданной на основе необъективных данных, может быть несправедливо по отношению к людям (или определенным группам людей). Справедливость связана с человеческой точкой зрения и основывается на личных и принятых в обществе нормах и убеждениях. Несправедливость в поведении системы ИИ может оказывать негативное, вредное и разрушительное воздействие на индивидуума или на группы людей. Дополнительную информацию о предвзятости систем ИИ см. в *ГОСТ Р 71476—2024* (пункт 5.15.9) и *[5]*.

5.5.3 Риски и управление рисками

5.5.3.1 Риски, как и степень доверия, являются нефункциональными свойствами систем ИИ. Системы ИИ, как традиционные системы программного обеспечения, функционируют, подвергаясь рискам (см. *[6]*). Риски определяются степенью потенциального воздействия, вызванного отказом системы или ее нештатным поведением и потенциально пострадавшими при этом индивидуумами или сообществом (см. *[2]*). Риски могут быть снижены с помощью мер управления рисками. *Степень управления рисками, предпринимаемая организацией, зависит от уровня совокупных рисков, с которыми организация готова сталкиваться в процессе своей деятельности.* В некоторых случаях, если число неблагоприятных факторов значительно, эта проблема может быть целью управления рисками на всех стадиях

жизненного цикла системы. В настоящей статье описаны элементы, связанные с рисками и управлением рисками (см. *[6]*).

5.5.3.2 Инфраструктура управления рисками, необходимая, чтобы «помочь организации интегрировать управление рисками в конкретные действия и функции», характерные для разработки, предоставления, предложения или использования систем ИИ, представлена в *[6]*. Процессы, связанные с видами деятельности и функциями, использующими ИИ, включают систематическое применение мер и методов, обработки информации для оценки, противодействия, регистрации и снижения рисков, подробно описанные в *[6]*.

5.5.4 Этические и общественные проблемы

5.5.4.1 Общие положения

С одной стороны, технологии ИИ потенциально могут принести огромные преимущества обществу, организациям и индивидам. С другой стороны, применение этой технологии ведет к возникновению потенциальных широкомасштабных этических и общественных проблем. Обычно этические проблемы связаны со сбором, обработкой и раскрытием личных данных, а также с необъективностью данных, которые используются в машинном обучении и обрабатываются непрозрачными алгоритмами, не обладающими свойством объяснимости (см. *[4]*).

5.5.4.2 Этические основы

Этические основы ИИ могут быть созданы на базе существующих этических концепций и теорий, таких как этика добродетели, утилитаризм, деонтология и другие (см. *[4]*). В настоящей статье описан подход к определению условий использования приложения на основе ИИ (см. *[4]*). Организации, планирующие разработку и использование ИИ ответственным образом, могут рассмотреть принятие разнообразных принципов ИИ (см. *[4]*). Ключевые темы, связанные с принципами ИИ: отслеживаемость, безопасность, прозрачность, объяснимость, справедливость и отсутствие дискриминации, контроль человека над технологией, профессиональная ответственность, приверженность общечеловеческим ценностям и правам человека, уважение международных норм поведения, вовлеченность общества в процессы развития, уважение закона, охрана окружающей среды, трудовые отношения (другие примеры построения общественно приемлемых систем ИИ см. в *[4]*).

5.5.4.3 Общественные проблемы

Использование технологии ИИ потенциально может воздействовать на многочисленные заинтересованные стороны среди общественности, помимо клиентов и пользователей. Эти заинтересованные стороны могут быть членами сообщества, где внедряется технология ИИ, или представителями будущих поколений, которые будут жить, испытывая воздействие ИИ на качество жизни в физической среде или в ходе трудовой деятельности. На организации, планирующей разработку и использование ИИ, лежит ответственность перед обществом, обязанность определить заинтересованные стороны и учитывать воздействие технологии (см. *ГОСТ Р ИСО 26000*).

5.5.4.4 Правовые нормы и проблемы

Регулирование технологии ИИ является новой областью. Правовые нормы для разработки, внедрения и использования систем на основе ИИ находятся в стадии развития и полностью не определены. В некоторых регионах введены правовые нормы, определяющие отдельные аспекты для технологий и приложений на основе ИИ (например, распознавание лиц правоохранительными органами), вносятся и обсуждаются многочисленные предложения по регулированию ИИ. В настоящий момент не существует скоординированных и последовательных правовых норм регулирования ИИ на уровне предметной области, региона, страны или на международном уровне.

6 Взгляды заинтересованных сторон и концепция основных подходов к приложению на основе ИИ

6.1 Общие положения

Основной контекст применения приложения на основе ИИ, описанный в 5.2, определяет концепцию основных подходов к приложению на основе ИИ, изложенную в данном разделе. Эти подходы могут быть использованы для ответа на вопрос: «Каковы характеристики и общий анализ приложения на основе ИИ?»

6.2 Взгляд с точки зрения заинтересованных сторон

Концепция основных подходов к приложению на основе ИИ, описанная в настоящем стандарте, включает взгляды разных групп среди заинтересованных сторон. Эти группы имеют разные взгляды на приложение на основе ИИ в зависимости от своих намерений и целей. На рисунке 2 они представлены на уровнях производителей, пользователей и воздействия. Граница организации проведена на рисунке 2, чтобы отделить разные заинтересованные стороны. Взгляды сторон сфокусированы на анализе, функциональных и нефункциональных характеристиках приложения на основе ИИ.

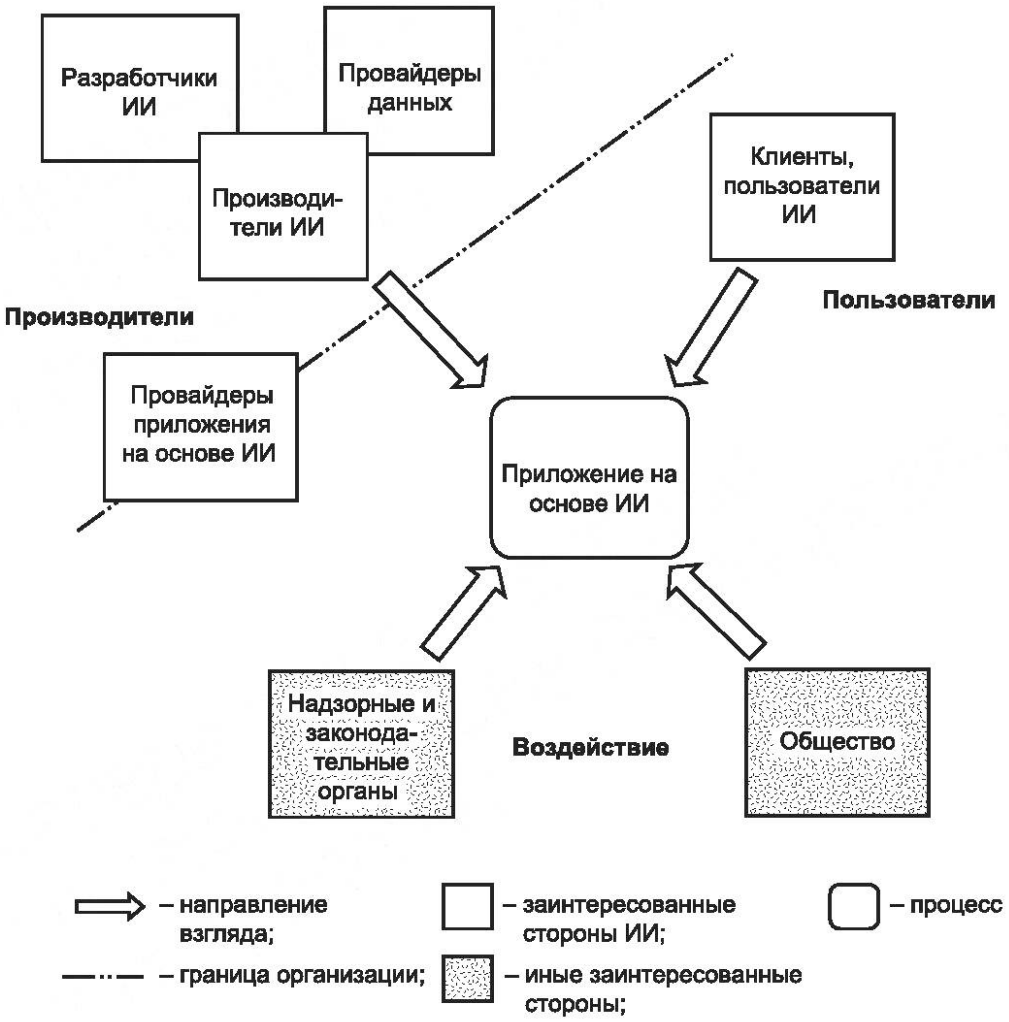

Рисунок 2 — Точки зрения заинтересованных сторон на приложение на основе ИИ

Взгляд производителя — это точка зрения производителей и разработчиков ИИ и поставщиков данных, которые создают приложение на основе ИИ. Провайдер ИИ разделяет до определенной степени эти взгляды, поскольку может играть некоторую роль и в производстве, и во внедрении приложения на основе ИИ. Клиент и пользователь ИИ смотрят с точки зрения использования, поскольку именно они применяют приложение на основе ИИ как дополнительный инструмент при принятии решения. Точка зрения общественности исходит из того, как на общество влияет ввод в действие приложения. Надзорные органы рассматривают то, насколько внедрение соответствует местным правовым нормам (за которые отвечают законодательные органы) и какие меры юридического воздействия применяются к продукту или сервису, если он не соответствует нормам.

6.3 Среда разработки приложения на основе ИИ

Комбинация взглядов заинтересованных сторон на среду разработки приложения на основе ИИ показана на рисунке 3.

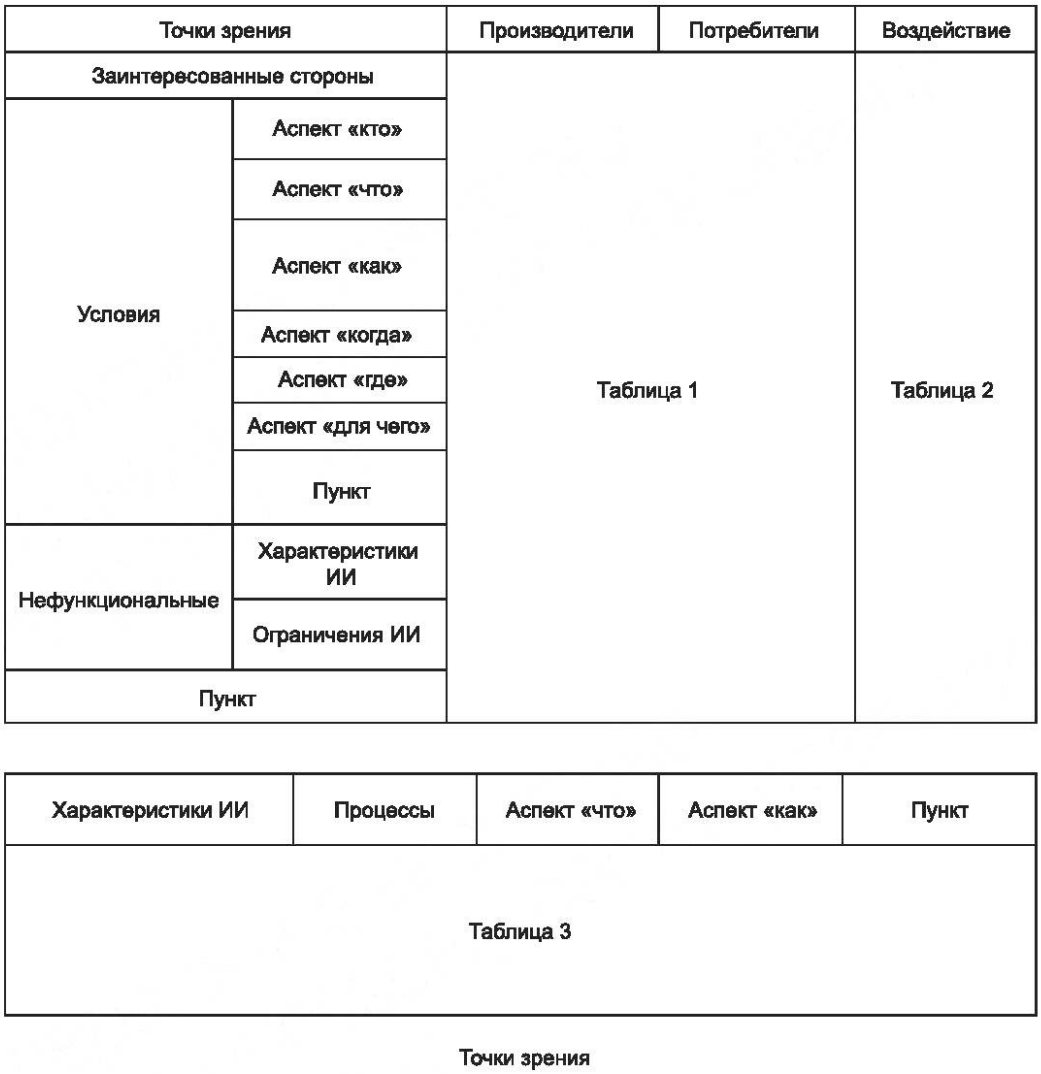

Рисунок 3 — Комбинация взглядов заинтересованных сторон на среду разработки приложения на основе ИИ

В таблице 1 приведены взгляды производителя и пользователя на среду разработки приложения на основе ИИ и взаимосвязь заинтересованных сторон, в таблице 2 — воздействие ИИ на заинтересованные стороны, в таблице 3 — характеристики ИИ и вовлеченность заинтересованных сторон в процессы.

Таблица 1 — Взгляды производителя и пользователя

Точки зрения		Производитель				Пользователь	
Заинтересованные стороны		Производители ИИ	Поставщики данных	Разработчики ИИ	Провайдеры приложения ИИ	Клиенты ИИ	Пользователи ИИ
Условия	Аспект «кто»	+	+	+	+	+	+
	Аспект «что»	+	+	+	+	+	+
	Аспект «как»	Создание, использование, обновление	Создание	Создание	Создание	Использование	Использование
	Аспект «когда»	Все стадии	Проектирование, разработка, валидация и верификация	Проектирование, разработка, валидация и верификация	Разработка, эксплуатация и мониторинг	Все стадии	Разработка, эксплуатация и мониторинг
	Аспект «где»	+	+	+	+	+	+
	Аспект «для чего»	+	+	+	+	+	+
	Пункты	7.2.2	7.2.3	7.2.4	7.2.5	7.3.2	7.3.2
Нефункциональные	характеристики ИИ	Доверие к ИИ, риски и управление ими					
	ограничения ИИ	Этические и общественные проблемы, юридические нормы и проблемы					
Пункты		5.5.2—5.5.4					
Примечание — Знаком «+» отмечено участие заинтересованных сторон в условиях применения приложения на основе ИИ.							

Таблица 2 — Воздействие искусственного интеллекта

Точки зрения		Воздействие	
Заинтересованные стороны		Общество	Надзорные и законодательные органы
Условия	Аспект «кто»	+	+
	Аспект «что»	+	+
	Аспект «как»	+	+
	Аспект «когда»	Ввод в действие, эксплуатация, мониторинг	Ввод в действие, эксплуатация, мониторинг
	Аспект «где»	+	+
	Аспект «для чего»	+	+
	Пункт	7.4.2	7.4.3
Нефункциональные	характеристики ИИ	Доверие к ИИ, риски и управление ими	
	ограничения ИИ	Этические и общественные проблемы, юридические нормы и проблемы	
Пункты		5.5.2—5.5.4	
Примечание — Знаком «+» отмечено участие заинтересованных сторон в условиях применения приложения на основе ИИ.			

Таблица 3 — Характеристики искусственного интеллекта

Характеристики ИИ	Процессы	Аспект «что»	Аспект «как»	Подраздел
Применяемая система ИИ использует модель для получения информации и обрабатывает ее с помощью человека или без нее, с помощью алгоритма или программы	Разработка модели ИИ, система ИИ	+	+	5.4
Применение оптимизации или логического вывода модели для улучшения решений, прогнозов или рекомендаций для конкретных целей	Применение ИИ, дополнительное принятие решения	+	+	
Обновление и совершенствование модели, системы или приложения при оценке результатов взаимодействия	Непрерывная валидация	+	+	
Примечание — Знаком «+» отмечены характеристики ИИ в условиях применения приложения на основе ИИ.				

7 Руководство для приложений на основе ИИ

7.1 Общие положения

Настоящий раздел содержит рекомендации заинтересованным сторонам для осознания ими своей роли и ответственности, а также понимания перспектив при создании, использовании приложений на основе ИИ или при ответной реакции на воздействие такого применения ИИ. В данном руководстве учтены функциональные (см. 5.4) и нефункциональные (см. 5.5) характеристики приложения на основе ИИ.

Руководство составлено в виде набора вопросов, которые нужно задать каждой заинтересованной стороне по поводу их роли и сферы ответственности, исходя из точек зрения, представленных на рисунке 2, среды разработки приложения на основе ИИ, приведенной в таблице 1, воздействия ИИ, приведенного в таблице 2, и характеристик ИИ, приведенных в таблице 3. Некоторые из представленных здесь вопросов могут дать возможность заинтересованным сторонам изучить глубже некоторые аспекты использования ИИ, такие, как применимость соответствующих международных стандартов к конкретному приложению. В приложении А приведены примеры ответов для заполнения таблиц с вопросами по конкретным практическим кейсам использования приложений на основе ИИ (см. таблицы А.1—А.3).

7.1.1 Общие положения о производителе

Те заинтересованные стороны, которые рассматривают вопрос с точки зрения производителя — это работники производителя ИИ или его партнеры, которые проектируют, создают, занимаются верификацией и валидацией системы ИИ и поставляют для нее данные. Систему ИИ, в свою очередь, вводят в действие в виде приложения.

7.1.2 Точка зрения производителя

Производителю ИИ следует принять во внимание как минимум следующие обстоятельства:

- кем являются клиенты и пользователи ИИ?
- кем являются разработчики ИИ? Являются ли они квалифицированными и опытными работниками, штатными или нанятыми по контракту?
- кем являются провайдеры приложения и в каких отношениях они с производителем ИИ?
- кем являются заинтересованные стороны на каждом этапе жизненного цикла системы ИИ [см. *ГОСТ Р 71476—2024* (раздел 6)]? Что представляет собой система ИИ и каковы ее возможности? Какой алгоритм используется для модели ИИ?
- каковы характеристики ИИ, используемые в данном конкретном приложении?
- какие данные использованы для создания модели ИИ? Каковы источники этих данных? Кем являются поставщики данных и их партнеры?
- какова степень доверия к приложению на основе ИИ и каковы связанные с ним риски? Что делается для оценки и снижения рисков? Существует ли в организации система управления рисками (см. *[2]*, *[6]*)?

- какие вопросы (этические, социальные проблемы, безопасность, конфиденциальность, неприкосновенность личных данных и другие вопросы, связанные с правовыми нормами) возникают в связи с созданием и вводом в действие приложения на основе ИИ? Каким образом эти вопросы решаются?

- какова система инструментальных средств оценки при вводе в действие приложения на основе ИИ?

- каково качество системы ИИ в целом (см. *[3]*, *[7]*)?

- каким образом приложение на основе ИИ создают, применяют и обновляют? Каким образом модель ИИ обучают или программируют? Насколько робастна модель ИИ (см. *ГОСТ Р 70462.1*)? Когда (на каком этапе жизненного цикла системы ИИ) структура модели, приложение и обновления будут анализировать и пересматривать? Где модель создают, применяют и обновляют, локально или с использованием служб облачных вычислений? Когда (на каком этапе жизненного цикла системы ИИ) следует привлечь производителя для повторной оценки характеристик ИИ в среде?

- где планируют развернуть приложение (локально или в виде службы облачных вычислений)? Где будут разрабатывать приложение? Где находятся разработчики? Где находятся источники данных? Где, в какой географической точке, будет развернуто приложение?

- по какой причине приложение на основе ИИ разрабатывают в виде продукта или в виде сервиса? Какова его потенциальная значимость для производителя ИИ и пользователя? Каковы перспективы использования разработки и план действий?

7.1.3 Точка зрения поставщика данных

Поставщику данных следует принять во внимание как минимум следующие обстоятельства:

- кем является производитель ИИ: нанимателем, партнером, клиентом?

- какие данные собирают и каков их источник? Каким образом данные собирают, хранят, обрабатывают, поставляют и вводят в модель ИИ (о данных для машинного обучения — см. *ГОСТ Р 71484.4*? Используют ли систему управления данными (см. *ГОСТ Р 71484.3*)?

- в какой области, географической и иной, собирают данные? Какие ограничения могут применять к модели, разработанной на основе этих данных?

- какие ограничения налагают с точки зрения источников и самой природы данных, отбираемых для обучения?

- каким образом измеряют и валидируют качество собранных данных (см. *ГОСТ Р 71484.2*)? Какие проблемы связаны со степенью доверия к данным и с обеспечением их непредвзятости? Какие меры принимают для оценки этих проблем и противодействию им (см. *[2]* и *[5]*)?

- каким образом собирают, валидируют и используют данные для обновления модели ИИ на стадии ее эксплуатации и сопровождения? Каким образом собранные данные защищают и используют в соответствии с внутренними правилами и требованиями суверенитета данных?

- когда (на каких стадиях жизненного цикла системы ИИ) доступ к данным и их качество требуют пересмотра?

- где находятся источник данных? Где данные будут обрабатывать, (локально или с использованием сервиса облачных вычислений)? В какой географической точке?

- почему именно эти конкретные данные требуются для условий использования данного приложения на основе ИИ?

7.1.4 Точка зрения разработчика ИИ

Разработчику ИИ следует принять во внимание как минимум следующие обстоятельства:

- кем являются пользователь ИИ, поставщик данных и производитель ИИ?

- каковы отношения между разработчиком и производителем ИИ? Производитель ИИ является штатным работником или работником по контракту?

- какой уровень квалификации и опыт требуется для разработчиков ИИ?

- какую модель ИИ используют, обучают или программируют? Каким образом модель ИИ проектируют, создают, валидируют и верифицируют в рамках функциональных характеристик системы ИИ? Какие процессы в этом задействованы?

- какие существуют технические и системные требования для ввода в действие системы ИИ в виде доступного приложения?

- какие алгоритмы используют для обработки данных? Каковы критерии качества данных? Каковы критерии качества информации на выходе? Каковы критерии валидации и верификации? Каковы критерии для обновления модели?

- каким образом данные обрабатывают на предварительном этапе? Как определяют их качество? Как происходит выбор алгоритма? Каким образом адаптируют требования к модели?

- когда (на каких стадиях жизненного цикла системы ИИ) оценивают условия ее использования и требования к системе?

- где приложение на основе ИИ может быть развернуто (локально или в виде сервиса облачных вычислений)?

- для чего приложение создают в виде продукта или сервиса? Почему используют конкретную модель?

7.1.5 Взгляд провайдера приложения

Провайдеру приложения на основе ИИ следует принять во внимание и рассмотреть как минимум следующие вопросы:

- кем являются клиенты и пользователи ИИ и как они используют приложение?

- каковы отношения между производителем ИИ и провайдером? Являются ли они отношениями с нанимателем или с партнером?

- каковы характеристики ИИ в данном приложении? Каковы его возможности, вычислительная мощность, производительность и ограничения?

- каковы технические и системные требования для клиентов и пользователей для получения доступа и применения приложения? Какие предусмотрены меры восстановления на случай отказа?

- каковы аналитические данные об эксплуатационном состоянии приложения и как эти данные отслеживают?

- какое воздействие оказывает приложение на основе ИИ на клиентов, пользователей и сообщество в целом?

- каким образом приложение на основе ИИ создают, применяют и обновляют? Каким образом модель ИИ обучают или программируют? Насколько робастна модель ИИ (см. *ГОСТ Р 70462.1*)? Когда (на каком этапе жизненного цикла системы ИИ) структура модели, приложение и обновления будут пересматривать и анализировать? Где модель создают, применяют и обновляют, локально или с использованием сервисов облачных вычислений? Когда (на каком этапе жизненного цикла системы ИИ) следует привлечь производителя для повторной оценки характеристик ИИ в среде?

- как происходит управление рисками при вводе в действие приложения?

- когда (на каком этапе жизненного цикла системы ИИ) оценивают рабочую среду и требования?

- применяют ли какие-то ограничения для использования приложения: рекомендуемое, допустимое и ответственное использование? Являются ли они юридическими нормами, включенными в лицензию программного обеспечения?

- где будет внедрено приложение? Какие нормы закона применяют к функциональным и нефункциональным характеристикам сферы применения приложения? Кем осуществляется надзор?

- для чего приложение создают в виде продукта или сервиса?

7.2 Взгляд с точки зрения использования

7.2.1 Общие положения

Заинтересованные стороны, рассматривающие вопрос с точки зрения пользователя — клиенты и пользователи ИИ, которые применяют приложение на основе ИИ как инструмент оптимизации процесса принятия решения (см. 5.3.4.6).

7.2.2 Точка зрения заказчика (клиента) и пользователя

Клиенту или пользователю ИИ следует, как минимум, рассмотреть:

- каковы отношения между провайдером приложения и клиентом или пользователем ИИ?

- каковы ограничения для клиента и пользователя ИИ (как представителей общественности) при использовании приложения? Каковы последствия применения приложения в организации с точки зрения политики управления (см. *[8]*)?

- какие данные собирают при применении приложения на основе ИИ и каким образом их используют (для машинного обучения см. *ГОСТ Р 71484.2*)? Какие приняты принципы управления данными? Вводят ли данные повторно в модель ИИ для непрерывного обучения и совершенствования?

- какова степень доверия к приложению на основе ИИ и каковы связанные с ним риски? Что делают для оценки и снижения рисков? Существует ли в организации система управления рисками (см *[2]*, *[6]*)?

- каковы прозрачность и объяснимость приложения, предложенного провайдером?

- каковы этические и общественные проблемы, связанные с использованием приложения на основе ИИ? Как их решают?

- принятие какого рода решений будет оптимизировано с помощью приложения? Каков уровень его автоматизации? Кто оценивает эффективность приложения и какие метрики используют?

- каким образом клиенты и пользователи получат доступ к выходным результатам приложения для использования в процессе принятия решения? Каким образом измеряют производительность и эффективность приложения?

- когда (на какой стадии жизненного цикла системы ИИ) оценивают рабочую среду (контекст) приложения и требования к приложению? Каковы правовые нормы при внедрении приложения?

- для каких целей применяют приложение на основе ИИ? Каковы потенциальные выгоды и преимущества от использования приложения?

7.3 Возможные последствия применения ИИ

7.3.1 Общие положения

Сообщество, в котором вводят в действие приложение на основе ИИ, и отдельно взятые потребители в этом сообществе могут испытать на себе воздействие от применения приложения. В качестве примера можно привести использование приложений для видеонаблюдения, при подаче заявок на кредиты, предоставлении медицинских услуг, распространении информации через соцсети и другие. На внедрение приложений могут влиять местные надзорные органы, имеющие юрисдикцию над использованием технологии ИИ на основе правовых норм, установленных законодательными органами.

7.3.2 Точка зрения общественности

Сообществу, в котором вводят в действие приложение на основе ИИ, необходимо принять во внимание и рассмотреть как минимум следующие вопросы:

- кем являются потребители? Что конкретно их беспокоит как членов сообщества?

- какие данные собирают, применяя приложение, и как их используют? В чем состоит проблема хранения личных данных?

- каким образом использование приложения на основе ИИ влияет на сообщество и потребителей? Как это воздействие измеряют, как часто и кто? Какими способами сообщество устраняет негативное воздействие ИИ?

- когда (на каком этапе жизненного цикла системы ИИ) оценивают и пересматривают обратную связь от клиентов или требования клиентов?

7.3.3 Точка зрения надзорных и законодательных органов

Надзорным и законодательным органам следует принять во внимание и рассмотреть как минимум следующие вопросы:

- кем являются потребители? Что конкретно их беспокоит как членов сообщества?

- каков механизм принятия правовых норм для ввода в действие приложений на основе ИИ (инициатива идет сверху или снизу)? Каким образом используют конкретное приложение и как это влияет на конкретное сообщество? Кто несет за это ответственность (например, провайдер, клиент, пользователь)?

- когда (на каком этапе жизненного цикла системы ИИ) правовые нормы оценивают и пересматривают?

- где внедряют приложение? Каковы применимые к нему правовые нормы? Каким образом планируют отслеживать соответствие конкретного приложения установленным нормам? Кто является ответчиком в случае нарушения законодательных норм?

- для каких целей применяют приложение на основе ИИ? Каковы потенциальные выгоды от использования приложения? Каково его потенциальное положительное или негативное воздействие на общество?

Приложение А
(справочное)

Примеры использования приложений на основе ИИ

А.1 Общие положения

В настоящем приложении приведено два примера для демонстрации применения настоящего стандарта с учетом условий использования (см. 5.2), заинтересованных сторон и процессов (см. 5.3), характеристик ИИ (см. 5.4), отраженных в общей концепции подходов для приложения на основе ИИ (см. раздел 6 и таблицу 3). Некоторые ячейки таблиц А.1—А.6 заполнены ответами на вопросы, приведенными в разделе 7.

А.2 Пример использования приложения на основе ИИ Фуджитсу Лимитед для обнаружения дефектов в лопастях ветровых турбин

В настоящем разделе приведен пример из практики компании Фуджитсу Лимитед для приложения на основе ИИ, в котором использовано глубокое обучение алгоритма для обнаружения дефектов в лопастях ветровых турбин с целью повышения качества контроля и принятия решений при осмотре. Заинтересованные стороны, условия применения и характеристики ИИ приведены в таблицах А.1—А.3 с учетом основ применения ИИ (см. 6.3).

Таблица А.1 — Взгляд производителя на примере Фуджитсу Лимитед

Точки зрения		Производитель			
Заинтересованные стороны		Производители ИИ	Поставщики данных	Разработчики ИИ	Провайдеры приложения ИИ
Контекст	Аспект «кто»	Фуджитсу Лимитед	Поставщик УЗ-сканеров	Фуджитсу Лимитед	Фуджитсу Лимитед
	Аспект «что»	Отвечает за всю систему	Предоставляет данные сканеров	Создает систему ИИ с обученной моделью	Получает систему ИИ и предлагает ее как приложение
	Аспект «как»	Создание, применение, обновление	Создание	Создание	Создание
	Аспект «когда»	На всех стадиях	Проектирование, разработка, валидация и верификация	Проектирование, разработка, валидация и верификация	Ввод в действие, эксплуатация, мониторинг
	Аспект «где»	—	Производитель	Производитель	Локация клиента
	Аспект «для чего»	Производитель должен создать качественное приложение для клиента, понимая цели и требования к приложению из диалога с клиентом	—	—	—

Таблица А.2 — Взгляд потребителя и воздействие ИИ на примере Фуджитсу Лимитед

Точки зрения		Пользователь		Воздействие	
Заинтересованные стороны		Клиенты ИИ	Пользователи ИИ	Общество	Надзорные и законодательные органы
Контекст	Аспект «кто»	Изготовитель ветротурбин	Специалист отдела технического контроля, инспектор изготовителя	Компания энергосбыта, пользователи и соседи ветротурбин	Например, в Европейском союзе Генеральный директорат по внутреннему рынку, промышленности, предпринимательству и малому и среднему предпринимательству
	Аспект «что»	—	—	—	—
	Аспект «как»	Применение	Применение	—	—
	Аспект «когда»	Ввод в действие, эксплуатация, мониторинг	Ввод в действие, эксплуатация, мониторинг	Ввод в действие, эксплуатация, мониторинг	Ввод в действие, эксплуатация, мониторинг
	Аспект «где»	Локация клиента	Локация клиента	—	—
	Аспект «для чего»	Дефекты лопастей в эксплуатации приводят к авариям и большому ущербу для репутации производителя. Производитель поставляет более 5000 лопастей турбин в год для ветроэлектростанций на суше и в море. Для лопасти до 75 м длиной требуется 6 ч работы опытного контролера для оценки данных сканирования в процессе контроля качества. Система ИИ сокращает время контроля на 80 %, снижая затраты, длительность производственного цикла, повышая производительность	Специалисты отдела технического контроля заинтересованы в повышении производительности путем применения ИИ и в точности работы приложения	Жители рядом с ветроэлектростанцией обеспокоены ущербом, наступление которого возможно при дефектах в ветровых турбинах	Надзорный орган гарантирует безопасность продукции на основе правовых норм регулирования производства крупноразмерных объектов, таких как лопасти турбин, предусмотрена ответственность за дефекты изделия

Таблица А.3 — Характеристики приложения на основе ИИ на примере Фуджитсу Лимитед

Характеристики ИИ	Процессы	Аспект «что»	Аспект «как»
Применяемая система ИИ использует модель для получения информации и обрабатывает ее с помощью человека или без нее, с помощью алгоритма или программы	Фаза обучения — алгоритмические процессы	Глубокое обучение и имиджификация	Преобразование исходных данных УЗ-сканирования в изображения с помощью основных цветов (красного, зеленого, синего)
Применение оптимизации или логического вывода модели для улучшения решений, прогнозов или рекомендаций для конкретных целей	Приложение — этап реализации	Поиск дефектов	Просмотр данных неразрушающего контроля качества с помощью УЗ-сканирования
	Принятие решений	Принятие решений на этапе контроля качества	Человек принимает решение, взаимодействуя с машиной. Система ИИ обнаруживает 95 % дефектов. Человек обследует только те части лопасти, которые выделены системой ИИ. Например, при отсутствии донного эхо-сигнала, наличии инородных включений, неровностей
Обновление и совершенствование модели, системы или приложения при оценке результатов взаимодействия	Непрерывная валидация	Глубокое обучение и имиджификация	Мониторинг результатов за определенный период после последнего или повторного обучения или до внедрения нового процесса производства

А.3 Пример использования приложения на основе ИИ в чат-ботах ЛайвПерсон для распознавания естественного языка

В настоящем разделе приведен пример из практики компании ЛайвПерсон для приложения на основе ИИ, в котором использовано машинное обучение для улучшения качества диалога машины с человеком с помощью программы, распознающей естественный язык, в сочетании с пониманием контекста и предметной области и анализом тональности высказывания. Заинтересованные стороны, контекст использования и характеристики ИИ приведены в таблицах А.4—А.6 с учетом основ применения ИИ (см. 6.3).

Таблица А.4 — Взгляд производителя на примере ЛайвПерсон

Точки зрения		Производитель			
Заинтересованные стороны		Производители ИИ	Поставщики данных	Разработчики ИИ	Провайдеры приложения ИИ
Контекст	Аспект «кто»	ЛайвПерсон	Клиент	ЛайвПерсон	ЛайвПерсон, провайдеры мессенджеров и обработки NLU[1)]

Окончание таблицы А.4

Точки зрения		Производитель			
Заинтересованные стороны		Производители ИИ	Поставщики данных	Разработчики ИИ	Провайдеры приложения ИИ
Контекст	Аспект «что»	Создание NLU, сетевого чата, набора средств разработки для мессенджеров, коннекторов для приложений (мессенджеров) для третьих лиц	Сбор транскрипты диалогов в чате и мессенджере между людьми и между ботами и людьми	Разработка механизма NLU, приложения и интерфейсов контактного центра управления участниками коммуникации	Предложение приложения для сообщений на базе операционных систем для ручных устройств, предложение приложения третьим лицам
	Аспект «как»	Создание, применение, обновление	Создание	Создание	Создание
	Аспект «когда»	На всех стадиях	Проектирование, разработка, валидация и верификация	Проектирование, разработка, валидация и верификация	Проектирование, разработка, валидация и верификация
	Аспект «где»	Частное облако	Частное облако и платформа мессенджера третьей стороны	Частное облако и платформа мессенджера третьей стороны	Частное облако и платформа мессенджера третьей стороны
	Аспект «для чего»	Привлекательность: дополняют возможности представителей клиентской службы и агентов в обслуживании клиентов. Задействуют чат-боты в тех случаях, когда представители и агенты заняты, или для фильтрации запросов с помощью анализа тональности, направляя к другим ботам или сотрудникам	NLU должна соответствовать требованиям клиентов и отрасли. Клиент может по желанию создать собственный бот и автоматизировать полностью или частично свои разговоры	Чат-коммерция требует специализированной языковой модели и анализа интенций. Хотя ЛайвПерсон может использовать NLU других провайдеров, собственная NLU разработана по запросам конкретных клиентов	—

[1)] NLU — Natural language understanding (обработка и понимание естественного языка).

Таблица А.5 — Взгляд потребителя и воздействие приложения на основе ИИ на примере ЛайвПерсон

Точки зрения		Пользователь		Воздействие	
Заинтересованные стороны		Клиенты ИИ	Пользователи ИИ	Общество	Надзорные и законодательные органы
Контекст	Аспект «кто»	Клиент	Пользователи сети	Члены сообщества	Для конкретной сферы, например страхование, банкинг и обработка кредитных карт, телекоммуникации и средства массовой информации, здравоохранение. В зависимости от применения, например нужно идентифицировать как автоматизированный сервис. В зависимости от требований конфиденциальности, например требуется для защиты данных клиентов
	Аспект «что»	Задают распределение сообщений от клиентов людям или ботам		—	—
	Аспект «как»	Применение	Применение	—	—
	Аспект «когда»	Ввод в действие, эксплуатация, мониторинг			
	Аспект «где»	Частное облако		—	—
	Аспект «для чего»	Крупные контактные центры и мессенджеры нуждаются в распределении обращений по обслуживанию, продажам, технической поддержке и жалобам групп специалистов		—	Многие отрасли регулируют для защиты потребителей и общества в целом. Сообщения имеют по сравнению с голосовым общением или личным, диалог ведется письменно, организации отслеживают его качество в реальном времени и видят риски, уязвимых клиентов и потенциально неприемлемое поведение

Таблица А.6 — Характеристики приложения на основе ИИ на примере ЛайвПерсон

Характеристики ИИ	Процессы	Аспект «что»	Аспект «как»
Применяемая система ИИ использует модель для получения информации и обрабатывает ее с помощью человека или без нее, с помощью алгоритма или программы	Машинное обучение, NLU	Конструктор разговоров помогает клиентам создавать чат-боты для собственных целей	Сбор данных для машинного обучения для совершенствования возможностей ИИ

Окончание таблицы А.6

Характеристики ИИ	Процессы	Аспект «что»	Аспект «как»
Применение оптимизации или логического вывода модели для улучшения решений, прогнозов или рекомендаций для конкретных целей	Чат-бот, анализ интонаций	Общение с потребителями посредством живого взаимодействия с представителями клиентской службы. Агенты и чат-боты интегрированы в NLU для анализа в реальном времени	Автономный веб-разговор с потребителями
	Агент службы спрашивает и отвечает	Дополнение возможностей агента анализом интонаций высказываний для выбора оптимальных ответов	Дополнение работы клиентской службы и агентов возможностями обслужить клиентов последовательно и эффективно
Обновление и совершенствование модели, системы или приложения при оценке результатов взаимодействия	Чат-боты, агенты службы	Переток данных обратно к модели	Мониторинг улучшений с обратной связью из диалога с клиентами, используя рейтингование для агентов и чат-ботов

На рисунке А.1 показаны компоненты приложения ЛайвПерсон на основе ИИ в его среде. В приложении используют интегрированные возможности обработки естественного языка и распознавания тональности высказывания для построения диалога в чатах с пользователями онлайн. Приложение также сохраняет данные этих диалогов для использования в машинном обучении как часть непрерывного совершенствования системы.

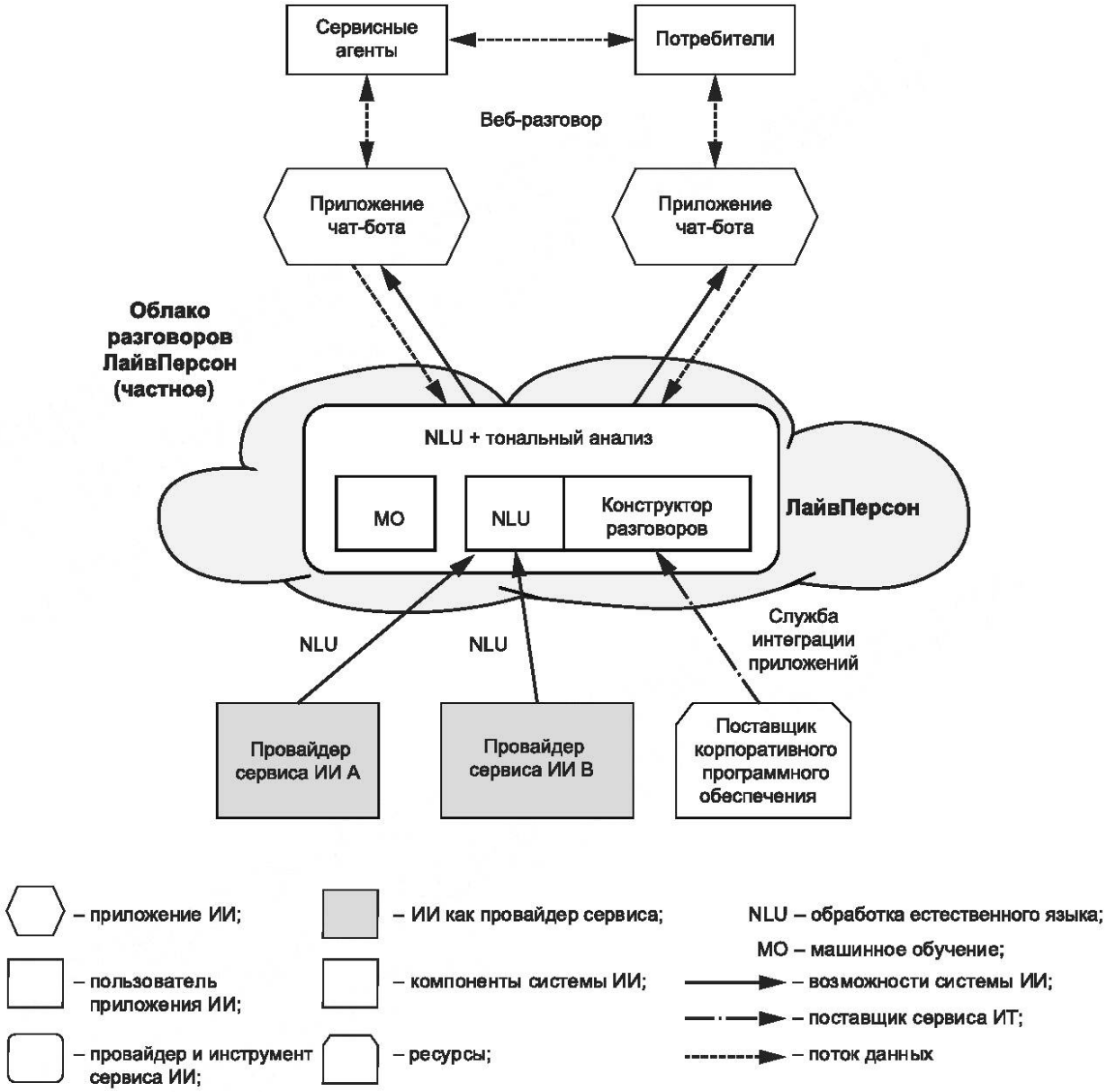

Рисунок А.1 — Компоненты приложения ЛайвПерсон

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 «Информационная технология. 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 23053:2022	Экосистема разработки систем искусственного интеллекта (ИИ) с использованием машинного обучения (МО) [Framework for Artificial Intelligence (AI) Systems using Machine Learning]
[2]	ISO/IEC TR 24028:2020	Информационная технология. Искусственный интеллект. Обзор вопросов доверенности в области искусственного интеллекта (Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence)
[3]	ИСО/МЭК 25059:2023	Программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модель качества для систем на основе искусственного интеллекта [Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems]
[4]	ISO/IEC TR 24368:2022	Информационные технологии. Искусственный интеллект. Обзор этических и социальных вопросов (Information technology — Artificial intelligence — Overview of ethical and societal concerns)
[5]	ISO/IEC TR 24027:2021	Информационные технологии. Искусственный интеллект (ИИ). Смещенность в системах ИИ и при принятии решений с помощью ИИ (Information technology — Artificial intelligence — Bias in AI systems and AI aided decision making)
[6]	ИСО/МЭК 23894:2023	Информационная технология. Искусственный интеллект. Руководство по менеджменту риска (Information technology — Artificial intelligence — Risk management)
[7]	*ISO/IEC DTS 25058*	*Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Руководство по оценке качества систем искусственного интеллекта [Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of artificial intelligence (AI) systems]*
[8]	ИСО/МЭК 38507:2022	Информационные технологии. Стратегическое управление ИТ. Последствия влияния стратегического управления при использовании искусственного интеллекта организациями (Information technology — Governance implications of the use of artificial intelligence by organizations)

УДК 004.01:006.354 ОКС 35.020

Ключевые слова: искусственный интеллект, руководство по применению искусственного интеллекта

Технический редактор *И.Е. Черепкова*
Корректор *О.В. Лазарева*
Компьютерная верстка *М.В. Малеевой*

Сдано в набор 28.10.2024. Подписано в печать 06.11.2024. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 3,72. Уч.-изд. л. 3,16.

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.
www.gostinfo.ru info@gostinfo.ru