



ПРЕДВАРИТЕЛЬНЫЙ
НАЦИОНАЛЬНЫЙ
СТАНДАРТ
РОССИЙСКОЙ
ФЕДЕРАЦИИ

**ПНСТ 841—
2023
(ISO/IEC DTS 25058)**

Системная и программная инженерия

**ТРЕБОВАНИЯ И ОЦЕНКА КАЧЕСТВА СИСТЕМ
И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ (SQuaRE)**

**Руководство по оценке качества
систем искусственного интеллекта**

(ISO/IEC DTS 25058, MOD)

Издание официальное

Москва
Российский институт стандартизации
2023

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 документа, 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 16 ноября 2023 г. № 61-пнст

4 Настоящий стандарт является модифицированным по отношению к международному документу ISO/IEC DTS 25058 «Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Руководство по оценке качества систем искусственного интеллекта» [ISO/IEC DTS 25058 «Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of AI systems», 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

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 4 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)

© Оформление. ФГБУ «Институт стандартизации», 2023

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

II

Содержание

1 Область применения	1
2 Нормативные ссылки	1
3 Термины и определения	2
4 Общие положения	2
5 Методика оценки качества	4
6 Функциональная пригодность	4
7 Уровень производительности	6
8 Совместимость	7
9 Удобство использования	7
10 Надежность	9
11 Защищенность	11
12 Сопровождаемость	11
13 Переносимость	12
14 Эффективность	13
15 Производительность	13
16 Удовлетворенность	13
17 Свобода от риска	13
18 Покрытие контекста	16
Приложение ДА (справочное) Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте	19
Библиография	20

Введение

Оценка системы искусственного интеллекта (ИИ) может представлять собой нетривиальную задачу. Кроме того, негативные эффекты от систем ИИ низкого качества весьма существенны, поскольку такие системы часто разрабатываются для упрощения автоматизации критически важных действий и решений, которые широко используются в промышленности, правительственных учреждениях и научных кругах.

Цель настоящего стандарта дать разработчикам руководство для надлежащей оценки качества систем ИИ. Настоящий стандарт не устанавливает точные величины параметров и пороговые значения, поскольку они варьируются в зависимости от характера конкретной системы. Вместо этого в настоящем стандарте приведены всесторонние указания, охватывающие существенные аспекты качества системы ИИ, для проведения его оценки.

Тестирование является всеобъемлющим с точки зрения верификации каждой характеристики и подхарактеристики (субхарактеристики) тестовыми стратегиями, но подробное описание методов тестирования и измерений представлено в других документах, например в [1].

ПРЕДВАРИТЕЛЬНЫЙ НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

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 AI systems

Срок действия — с 2024—01—01
до 2027—01—01

1 Область применения

Настоящий стандарт устанавливает руководство по оценке качества систем ИИ на основе использования моделей качества систем ИИ.

Настоящий стандарт применим во всех организациях, разрабатывающих и использующих ИИ.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р ИСО/МЭК 25022—2021 Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Измерение качества при использовании

ГОСТ Р ИСО/МЭК 25023—2021 Системная и программная инженерия. Требования и оценка качества систем и программной продукции (SQuaRE). Измерения качества системы и программной продукции

ГОСТ Р ИСО/МЭК 25040 Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Процесс оценки

ГОСТ Р МЭК 61508-2 Функциональная безопасность систем электрических, электронных, программируемых электронных, связанных с безопасностью. Часть 2. Требования к системам

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

ПНСТ 835—2023 (ISO/IEC TS 4213:2022) Искусственный интеллект. Оценка эффективности моделей и алгоритмов машинного обучения в задаче классификации

ПНСТ 837—2023/ISO/IEC TS 8200 Искусственный интеллект. Управляемость автоматизированных систем искусственного интеллекта

ПНСТ 838—2023 (ИСО/МЭК 23053:2022) Искусственный интеллект. Структура описания систем искусственного интеллекта, использующих машинное обучение

ПНСТ 839—2023 (ISO/IEC TR 24027:2021) Искусственный интеллект. Смещенность в системах искусственного интеллекта и при принятии решений с помощью искусственного интеллекта

ПНСТ 840—2023 (ISO/IEC TR 24368:2022) Искусственный интеллект. Обзор этических и общественных аспектов

ПНСТ 842—2023 (ИСО/МЭК 25059:2023) Программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модель качества для систем искусственного интеллекта

ПНСТ 843—2023 (ИСО/МЭК 38507:2022) Информационные технологии. Стратегическое управление информационными технологиями. Последствия влияния стратегического управления при использовании искусственного интеллекта организациями

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 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.

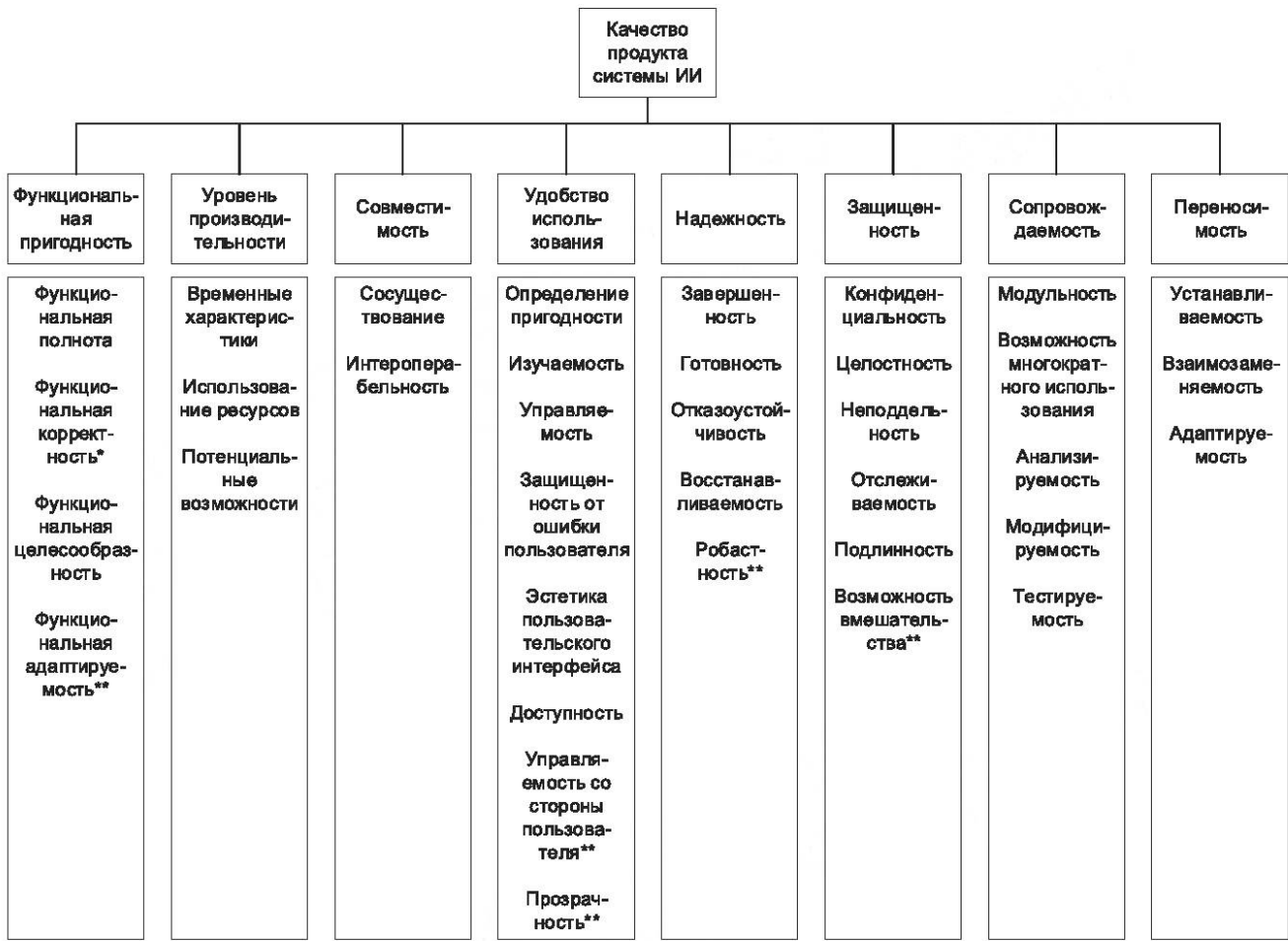
3 Термины и определения

В настоящем стандарте применены термины по ПНСТ 835—2023, ПНСТ 838—2023, ПНСТ 842—2023, [2], [3] и [4].

4 Общие положения

Чтобы учесть все релевантные аспекты качества системы ИИ в настоящем руководстве, даются ссылки на характеристики, определяемые в серии SQuaRE для модели качества продукта и модели качества при использовании для системы ИИ (см. ПНСТ 842—2023). Модель качества продукта и модель качества при использовании, применимые к любой системе, могут применяться и к системе ИИ. Несколько подхарактеристик были добавлены, некоторые подхарактеристики изменили свои значения или контексты.

На рисунках 1 и 2 (см. ПНСТ 842—2023) представлены характеристики и подхарактеристики модели качества продукта и модели качества при использовании для системы ИИ. Подхарактеристики добавлены или модифицированы по сравнению с подхарактеристиками стандарта модели качества продукта и модели качества при использовании для систем, поскольку системы ИИ отличаются от систем и программных средств в целом.



*Измененные подхарактеристики.

**Новые подхарактеристики.

Рисунок 1 — Модель качества продукта системы ИИ



*Новые подхарактеристики.

Рисунок 2 — Модель качества при использовании системы ИИ

5 Методика оценки качества

Руководство по оценке качества базируется на релевантных подхарактеристиках модели оценки качества.

В настоящем стандарте приведены все подхарактеристики модели оценки качества и модели оценки качества при использовании продукта серии стандартов SQaRE.

Настоящий стандарт дополняет процесс оценки качества, описанный в *ГОСТ Р ИСО/МЭК 25040* для систем ИИ.

6 Функциональная пригодность

6.1 Функциональная полнота

Качество функциональной полноты измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.2.1).

6.2 Функциональная корректность

Качество функциональной корректности измеряется в соответствии с показателями по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.2.2).

Функциональная корректность оценивается с помощью соответствующих ключевых показателей эффективности (key performance indicators, KPI) и измерений:

- следует установить показатели измерений и показатели эффективности для измерения способности системы выполнить конкретную задачу и для оценки степени непредсказуемости конкретной системы;

- следует использовать релевантные оценочные измерения функциональной корректности в зависимости от задач системы ИИ и целей заинтересованных сторон. Перечень типовых оценок приведен в *ПНСТ 838—2023*.

Функциональная корректность оценивается также с помощью следующих методов функционального тестирования:

- метаморфного тестирования: установление соотношения данных на входе и данных, полученных на выходе из системы;
- экспертной оценки: используется в том случае, когда система ИИ создается для замены экспертных оценок, метод базируется на привлечении группы экспертов для оценивания результатов тестирования;
- эталонной оценки системы ИИ (бенчмаркинг): используется в том случае, если система ИИ создается для замены существующей технологии, или в том случае, когда исходная система ИИ может быть использована в качестве эталона;
- тестирования поведения системы ИИ для разнообразных сценариев и условий, предлагаемых заинтересованными сторонами;
- тестирования в симулированной окружающей среде: используется в том случае, когда система ИИ оказывает физическое воздействие на окружающую среду;
- тестирования системы на месте использования: используется в том случае, если существуют потенциальные различия или изменения условий функционирования в реальных условиях эксплуатации по сравнению с условиями тестовой среды;
- управления рисками: тестирование поведения систем ИИ в соответствии с выявленными сценариями риска.

Оценка функциональной корректности должна проводиться на основе разных и репрезентативных наборов данных:

- следует выбирать наилучшую модель машинного обучения (МО), применяя соответствующие оценочные измерения к набору данных для валидации. Простая валидация использует один отложенный набор данных, однако предлагается, при возможности, использовать кросс-валидацию на K-наборах данных;
- на стадии отдельно проводимого бэк-теста выбранную модель МО следует протестировать еще раз на базе нового (тестового) набора данных для проверки ее состоятельности;
- наборы данных для обучения, валидации и тестирования должны создаваться на основе разнообразных данных;
- наборы данных для валидации и тестирования должны содержать репрезентативные выборки, отражающие реальные условия;
- модель МО должна быть проверена на наборах данных с известными когортами, чтобы определить положительное или отрицательное изменение смещения.

Окончательные значения в настройках модели МО (например, пороговые значения при классификации) устанавливаются в сотрудничестве с пользователями.

Функциональная корректность для целей мониторинга должна оцениваться на реальных данных. Система вводится в эксплуатацию после успешного прохождения ею бэк-теста.

6.3 Функциональная целесообразность

Качество функциональной целесообразности измеряется в соответствии с показателями функциональной пригодности по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.2.3).

6.4 Функциональная адаптируемость

Система ИИ должна иметь механизм динамической адаптации к изменяющимся данным производственного (технологического) процесса, для чего используются следующие методы:

- метод непрерывного обучения или обучения с подкреплением;
- метод автоматического пайплайна по обновлению параметров модели (автокалибровка).

Примечание — Функциональная адаптируемость не подразумевает обязательного изменения самих целей системы, поскольку такие изменения могут привести к трансформации ее функционального состояния.

Построить систему адаптации с петлей обратной связи. Эта управляемая система включает четыре основные функции: мониторинг — анализ — планирование — исполнение:

- функция мониторинга связана с отслеживанием управляемой системы и среды, в которой система функционирует, и с обновлением информации;
- функция анализа связана с использованием обновленной информации для оценки потребности в адаптации, используя методы всестороннего анализа или моделирование рабочего цикла моделями МО;

- функция планирования связана с выбором наилучшего варианта в зависимости от целей адаптации и разработкой плана по адаптации существующей конфигурации системы и преобразованию ее в новую конфигурацию;

- функция исполнения связана с выполнением адаптационных действий.

Функциональная адаптируемость оценивается с использованием измерений, ключевых показателей эффективности (KPI) и методов функционального тестирования, описанных в 6.2 для измерения адаптируемости системы ИИ к новому набору данных.

Следует принимать во внимание необходимость компромиссных решений в отношении затратности при выборе наилучшей модели МО, поскольку модели, дающие наиболее точные прогнозы, могут оказаться непомерно дорогостоящими с точки зрения затрат на вычисления при оценивании (см. 7.2).

7 Уровень производительности

7.1 Временные характеристики

Качество временных характеристик измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.3.1).

Следует определять временные характеристики во время циклов обучения, оценки качества и исполнения при нормальных условиях как часть нормальных рабочих циклов в производственном процессе, в производственной среде, с использованием производственной инфраструктуры и вычислительных ресурсов, поскольку временные характеристики зависят от используемых ресурсов (см. 7.2).

При измерении длительности процесса следует принимать во внимание механизм адаптируемости системы ИИ. Например, в системе, включающей последовательность повторной настройки, оценивания и исполнения, должна учитываться длительность всей последовательности рабочих циклов в совокупности.

Следует протестировать и определить потенциальные конфликты между вычислительными ресурсами. Например, если в циклах обучения и исполнения используются одни и те же вычислительные ресурсы или исполнение проводится на многих экземплярах одновременно, это может отрицательно повлиять на временные характеристики системы.

Следует тестировать и определить распределение времени между сбором данных, их преобразованием и другими рабочими циклами, связанными с обработкой данных, в системе ИИ. Например, система не может выдать результат, если необходимые для этого входные данные предварительно не собраны и не преобразованы надлежащим образом.

7.2 Использование ресурсов

Качество подхарактеристики использования ресурсов измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.3.2).

Следует заложить достаточные ресурсы для процесса обучения системы ИИ на базе МО. Такие факторы, как типы вычислительных ресурсов, временные ограничения, количество данных, тип модели МО и число гиперпараметров, могут влиять на то, сколько ресурсов потребуется для завершения обучения.

Система ИИ может получать данные от других систем ИИ, что следует учитывать, поскольку это может создать дополнительную зависимость от используемых ресурсов.

Следует заложить необходимые ресурсы в целях реализации стратегии адаптируемости системы (использование пошагового обучения, активного обучения, онлайн-обучения или автоматическая калибровка). Например, стратегия адаптируемости, предусматривающая повторное переобучение или перекалибровку каждый час, требует гораздо больше ресурсов на постоянное переобучение по сравнению с этапом исполнения модели. Отдельного внимания заслуживают подходы, предусматривающие одновременность исполнения и дообучения модели.

7.3 Потенциальные возможности

Качество потенциальных возможностей измеряется в соответствии с показателями пропускной способности по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.3.3).

8 Совместимость

8.1 Сосуществование

Качество сосуществования измеряется в соответствии с показателями совмещения по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.4.1).

8.2 Интероперабельность

Качество интероперабельности измеряется в соответствии с показателями функциональной совместимости по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.4.2).

9 Удобство использования

9.1 Определимость пригодности

Качество определимости пригодности измеряется в соответствии с показателями полноты описания по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.1).

9.2 Изучаемость

Качество изучаемости измеряется в соответствии с показателями обучаемости по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.2).

Кроме того, измерения, обычно используемые для оценки изучаемости системы ИИ, включают (но не ограничиваются) следующие параметры:

- выполнение задачи: например, обучение вводу требуемых переменных через интерфейс пользователя для получения от системы ИИ прогноза с помощью модели МО. Измерения включают оценку способности выполнить задачу за установленное время, процентную долю пользователей, завершивших выполнение задачи, характеристики пользователей, которые завершили выполнение задачи, и другие данные;
- использование: например, обучение пользователя командам, необходимым для управления системой ИИ. Измерения оценивают успешность использования команд, число команд, используемых за определенное время, и другие показатели;
- когнитивные процессы: например, время, которое требуется пользователю для осознания и понимания вывода, полученного с помощью модели МО;
- полученная от пользователя обратная связь: мнение пользователя после обучения о желании пользоваться системой ИИ. Например, содержит ли система какие-либо противоречащие интуиции параметры или функции?

Характеристики изучаемости зависят от способностей и опыта конкретного тестирующего или пользователя. Например, пользователи с техническим образованием или те, кто имеет общее представление о системах ИИ, могут обучиться выполнению задач и использованию команд значительно быстрее. Пользователям без технического образования потребуется хороший пользовательский интерфейс для успешного применения системы ИИ. Чтобы точно оценить изучаемость, важно собрать смешанную группу пользователей разного уровня.

Оценка показателей изучаемости может осуществляться (но не сводится к ним) следующими методами:

- тестирования пользователем: пользователи могут оценить степень своей обученности использованию системы ИИ. Например, понятно ли им руководство пользователя? Понятны ли переменные, которые требуются для работы с конкретной системой ИИ? Насколько полно поля по умолчанию соответствуют реальным сценариям? Соответствуют ли данные на выходе системы ИИ выходным данным, описанным в руководстве пользователя?
- других инструментов для оценки изучаемости: кривая обучаемости, которая отражает соотношение времени и количества повторений выполнения задачи. Она показывает: а) обучаемость при первых попытках; б) нарастание обученности пользователей при повторном выполнении задачи (крутизну кривой); нарастание продуктивности пользователя при надлежащем освоении системы (выход кривой на конечный пологий участок).

9.3 Управляемость

Качество управляемости измеряется в соответствии с показателями эксплуатационной пригодности по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.3).

Управляемость системы ИИ может измеряться следующими методами:

- независимой экспертной оценки: независимые технические специалисты дают оценку системе ИИ на основе собственного опыта работы с системой. Например, специалист по МО может обнаружить, что временной интервал между процессом обработки данных и получением результата пользователем может быть сокращен. Специалист может предложить новые возможности для более эффективного развертывания модели;
- тестирования пользователем: пользователям можно предложить дать оценку степени соответствия системы ИИ их ожиданиям с точки зрения эффективности и точности. Например, насколько точно модель МО классифицирует объекты? Насколько сравнима такая классификация с выполненной человеком?
- иными методами оценки, связанными с характеристиками возможности многократного использования и надежности: стабильности системы ИИ в различных средах и с разными пользователями поможет достичь соответствия ожиданиям конкретного пользователя. Например, измерение времени обработки данных для получения результата в условиях применения разных переменных, вычислительные мощности и другие.

9.4 Защищенность от ошибки пользователя

Качество защищенности от ошибки пользователя измеряется в соответствии с показателями защиты от пользовательских ошибок по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.4).

9.5 Эстетика пользовательского интерфейса

Качество эстетики пользовательского интерфейса измеряется в соответствии с показателями эстетичности пользовательского интерфейса по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.5).

9.6 Доступность

Качество доступности измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.5.6).

9.7 Управляемость со стороны пользователя

Следует использовать параметры, указанные в *ПНСТ 837—2023*, для разработки системы, управляемой со стороны пользователя. Для оценки управляемости со стороны пользователя необходимо следующее:

- система ИИ в рабочем состоянии, с функционалом управляемости со стороны пользователя, соответствующим требованиям;
- тестовые элементы, разработанные специально для тестирования функционала управляемости со стороны пользователя;
- инструменты, способные выдавать команды, связанные с управляемостью, отслеживать внешнее состояние системы или внутренние изменения параметров, при этом просчитывая связанные показатели (например, задержка отклика и показатели стабильности).

Цели при оценке управляемости системы ИИ и соответствующий набор тестовых элементов должны быть следующими:

- следует тестировать соответствие функционала управляемости системы ИИ требованиям в целом, а также каждый элемент требуемого функционала отдельно, в сочетании с другими элементами;
- следует определять степень управляемости указанного функционала системы ИИ на том уровне, для которого система была разработана, создана или предназначена.

Управляющие подпроцессы, такие как передача управления, начало и конец управляющего воздействия, решение проблемы неопределенности при передаче управления, так же как и управление как таковое, должны быть протестированы. Для каждой составляющей управляющего функционала проверяется следующее (заинтересованные стороны могут применить дополнительный список в соответствии с конкретными запросами):

- корректность функционала управления: система ИИ может обучаться и выполнять задачи в соответствии с ожиданиями пользователя при получении корректной управляющей команды, остается в неизменном состоянии при получении некорректной управляющей команды;

- длительность управляющей функции: отрезок времени, который требуется, чтобы один из вспомогательных процессов или весь процесс управления в целом завершился. Длительность важных вспомогательных процессов (например, передача управления или начало управляющего воздействия) должна быть протестирована для системы ИИ, функционирующей в опасной среде;

- надежность управляющего функционала: степень стабильности, особенно в тех случаях, когда происходит отказ системы или непредвиденные инциденты. Для тестирования применяется методика внесения неисправностей;

- количество операций, требуемое для управляющего функционала: общее количество операций, которое должен выполнить контроллер для полного завершения управляющего процесса и получения результата. Учитываются все операции для всех вспомогательных процессов.

Более подробную информацию см. в ПНСТ 837—2023.

9.8 Прозрачность

Информация должна быть представлена заинтересованным лицам в открытой, всеобъемлющей и понятной форме.

Организация должна иметь возможность понять, отследить и документировать все обстоятельства, связанные с конфиденциальностью обработки данных, включая правовые, технические и организационные аспекты.

Система ИИ должна иметь прозрачных владельцев, ответственных за получение ожидаемой прибыли и информирующих заинтересованных лиц о результатах функционирования системы.

Релевантные характеристики, используемые в системе ИИ, должны быть целостными, доступными, ясными и понятными заинтересованным лицам. Например, если в системе ИИ в качестве входящей информации используется строго фиксированное число переменных, это ограничение должно быть донесено в понятной форме, чтобы избежать недоразумений.

Следует донести риски, связанные с информацией, полученной системой ИИ на выходе (например, прогнозы, решения и действия), поскольку они могут оказывать влияние на общество, экономику или окружающую среду. Информация должна быть донесена в понятной форме, точно и своевременно, несмещенно и полно.

Следует сообщать полученные системой ИИ результаты (например, прогнозы, решения и действия) соответствующим заинтересованным лицам в полной, доступной и понятной форме.

Необходимая информация о системе ИИ и степени ее качества должна быть открыта соответствующим заинтересованным лицам.

10 Надежность

10.1 Завершенность

Качество завершенности измеряется в соответствии с показателями стабильности по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.6.1).

10.2 Готовность

Качество готовности измеряется в соответствии с показателями готовности к работе по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.6.2).

10.3 Отказоустойчивость

Качество подхарактеристики отказоустойчивости измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.6.3).

10.4 Восстанавливаемость

Качество подхарактеристики восстанавливаемости измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.6.4).

10.5 Робастность

Робастность системы ИИ следует измерять в нормальных условиях эксплуатации. Эти условия зависят от сферы применения системы ИИ (см. [5]). Ограниченная область должна быть:

- определена набором атрибутов, которые четко определены;

- достаточной для выполнения системой ИИ одной или более установленных задач в соответствии с назначением;

- данные, используемые для обучения, должны быть репрезентативными, имея характеристики те же, что и данные, которые предстоит обрабатывать при выполнении реальных задач.

Следующие действия предлагаются для измерения робастности системы ИИ в пределах ограниченной области:

- оценка наборов данных для обучения, валидации и тестирования с точки зрения полноты соответствия нормальным условиям эксплуатации;

- разработка конкретных сценариев для тестирования функционирования системы в широком диапазоне условий нормальной эксплуатации;

- использование симуляционных подходов для генерации тестовых данных (см. [6]), чтобы они соответствовали режиму эксплуатации в полном объеме;

- использование регуляризации, аугментации данных и случайных помех для повышения робастности системы ИИ при нормальных условиях эксплуатации;

- оценка функциональной корректности (см. 6.2, в котором представлено рекомендуемое руководство по функциональной корректности, а также *ГОСТ Р 70462.1*, в котором представлены показатели робастности).

Система ИИ должна быть устойчива к аномальным событиям и явлениям типа «длинный хвост» и «черный лебедь». Рекомендуются к применению следующие меры:

- используйте сценарии тестирования «длинный хвост» и «черный лебедь», тестирование с эталонной моделью и стресс-тесты;

- идентифицируйте пределы робастности системы ИИ;

- окончательное решение о том, достаточна ли робастность системы ИИ для ее последующего использования, принимается заинтересованными лицами на основе предоставленных им результатов анализа.

Система ИИ должна быть устойчивой к значительным неожиданным попыткам нарушения защиты. Следующие меры следует рассмотреть:

- исправление отказов и ошибок в соответствии с наилучшими практиками, например путем усиления защиты, тестирования и верификации стабильности системы ИИ методами метаморфного тестирования, аугментации данных, генеративно-сопоставительных сетей, тренировки для обработки атак, генерация уязвимых элементов во входных данных или обнаружение атак;

- применение специфических мер противодействия в области МО, например обнаружение отклонений от нормального состояния на входе и выходе системы ИИ и пометка новых злоупотреблений для дальнейшего снижения функциональных рисков.

Система ИИ должна генерировать репрезентативные выводы при отклонении от нормы в окружающей среде:

- использовать доверительную оценку или доверительный интервал для принятия решения об использовании сгенерированного вывода или инициирования резервного цикла обработки исключений;

- калибровать доверительную оценку или доверительный интервал таким образом, чтобы они репрезентировали неопределенность относительно правильности данных на выходе, для чего используются тесты на корректность калибровки на основе данных тестирования;

- интегрировать резервные инструменты, такие как разбор входящего потока вручную, набор простых эвристических правил, статистическая модель или специальная отдельная система ИИ, которая перезаписывает выводы основной системы, которые определяются как аномальные или имеют слишком высокую степень неопределенности исходя из доверительных оценок или доверительных интервалов;

- резервные инструменты должны быть робастными и должны функционировать в том случае, когда происходит сбой генерации вывода в системе ИИ.

Аппаратное оборудование системы ИИ должно быть робастным по отношению к типичным причинам отказов. С точки зрения таких причин не существует различий между обычным аппаратным оборудованием и тем, которое используется в системах ИИ. Список релевантных причин сбоев представлен в *ГОСТ Р МЭК 61508-2* и [7].

Система ИИ должна адаптироваться к изменениям окружающей среды (см. 6.4, в котором рекомендовано руководство по функциональной адаптивности).

11 Защищенность

11.1 Конфиденциальность

Качество конфиденциальности измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.7.1).

Следует оценивать процессы обработки данных системой ИИ на предмет обеспечения их неприкосновенности и конфиденциальности как на входе, так и на выходе, определяя возможность применения обратного инжиниринга данных, таким образом идентифицируя угрозу неприкосновенности личных данных и их целостности.

Следует создать план снижения таких рисков, включающий деперсонализацию, использование синтетических данных, снижение размерности или иные механизмы, в случае обнаружения риска утраты конфиденциальности.

11.2 Целостность

Качество целостности измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.7.2).

Следует оценить систему ИИ на предмет возможности нарушения ее целостности при вводе неверных или вредоносных данных.

Следует создать план снижения рисков, включающий обнаружение мошеннических операций и аномалий и другие механизмы, которые используются при выявлении угрозы нарушения целостности.

11.3 Неподдельность

Качество неподдельности измеряется в соответствии с показателями неопровержимости по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.7.3).

11.4 Отслеживаемость

Качество отслеживаемости измеряется в соответствии с показателями отчетности по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.7.4).

11.5 Подлинность

Качество подлинности измеряется в соответствии с показателями аутентификации по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.7.5).

11.6 Возможность вмешательства

Невозможность вмешательства является одним из критериев управляемости в соответствии с *ПНСТ 837—2023*.

12 Сопровождаемость

12.1 Модульность

Качество модульности измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.8.1).

12.2 Возможность многократного использования

Качество возможности многократного использования измеряется в соответствии с показателями повторного использования по *ГОСТ Р ИСО/МЭК 25023—2021* (пункт 8.8.2).

Следует сохранять и извлекать знания системы ИИ в процессе ее создания для последующего использования в соответствии с *ПНСТ 838—2023*. Система состоит из устройства входа, устройства применения параметров модели и устройства выхода. Система ИИ встраивает знания в свои веса после обучения, но конкретное значение этих весов требует контекстуализации как на входе, так и на выходе. Знания оцениваются с помощью устройства выхода.

Система ИИ, предусматривающая многократное использование, должна разрабатываться таким образом, чтобы внедренная параметризация модели не применялась в концептуальной области, которая не покрывается данными, использованными для обучения. Для такой системы требуется также

канал передачи данных, где различаются версии обучающих данных. Для изменения целевой области применения системы ИИ могут использоваться следующие подходы:

- повторное обучение: система ИИ, основанная на МО, полностью обучается заново с использованием набора обучающих данных, соответствующих новой концептуальной области. Этот подход к управлению целевой областью при использовании системы ИИ можно назвать «силовым»;
- перенос обучения (transfer learning): в системе ИИ, основанной на МО, лишь часть весов модели МО настраивают с использованием небольшого набора данных из новой концептуальной области, чтобы использовать веса, уже существующие в системе, для идентификации нужных признаков;
- рефрейминг: отдельные компоненты системы изменяются, что позволяет системе обрабатывать информацию в незнакомой концептуальной области. Этот подход включает настройку встроенных компонентов на входе, нормализацию или настройку компонентов на выходе или полное отключение компонентов системы;
- создание новой версии: основные функции системы ИИ расширяются таким образом, чтобы они включали и новую концептуальную область. Такой подход применим скорее к более простым системам, чем к тем, которые имеют сложную архитектуру, например нейросети.

Система ИИ не должна применяться для тех целей, для которых она не разрабатывалась, без специальных мер, которые обеспечивают ее гибкость. Такие меры могут включать дополнительное совершенствование признаков, повторную настройку и повторную валидацию для адаптации системы к новой концептуальной области. Система, позволяющая многократное использование, должна разрабатываться таким образом, чтобы допускать гибкость в отношении интерпретации новой информации при несущественной ручной настройке разработчиками. Для этого при первоначальном вводе системы в эксплуатацию требуется детально изучить то, каким образом информация репрезентируется внутри системы.

Валидация способности системы использоваться в новой концептуальной области должна проводиться с учетом руководства по функциональной корректности и ее параметров (см. 6.2).

12.3 Анализируемость

Качество анализируемости измеряется в соответствии с показателями подверженности анализу по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.8.3).

Следует использовать механизм для протоколирования прогнозов и объяснений системы ИИ.

12.4 Модифицируемость

Качество модифицируемости измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.8.4).

Следует проектировать механизмы адаптируемости таким образом, чтобы адаптация модели МО не приводила к возникновению дефектов. Разработчикам следует обеспечить повышение качества или, как минимум, его сохранение на прежнем уровне для каждой новой версии модели МО или всей системы. Если повторное обучение ведет к снижению качества, для его поддержки должны быть разработаны последовательности операций, выполнение которых инициируется при соответствующих условиях.

Руководство по функциональной адаптируемости приведено в 6.4.

12.5 Тестируемость

Качество тестируемости измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.8.5).

Следует применять всестороннюю и эффективную стратегию тестирования после каждого обновления системы, чтобы предотвратить возникновение дефектов.

Следует применять инструменты всестороннего, непрерывного мониторинга, направленного на профилактику дефектов.

13 Переносимость

13.1 Адаптируемость

Качество адаптируемости измеряется в соответствии с показателями адаптивности по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.9.1).

13.2 Устанавливаемость

Качество устанавливаемости измеряется в соответствии с показателями установки по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.9.2).

13.3 Взаимозаменяемость

Качество взаимозаменяемости измеряется в соответствии с показателями заменяемости по ГОСТ Р ИСО/МЭК 25023—2021 (пункт 8.9.3).

14 Эффективность

14.1 Эффективность

Качество эффективности измеряется в соответствии с показателями результативности по ГОСТ Р ИСО/МЭК 25022—2021 (подраздел 8.2).

15 Производительность

15.1 Производительность

Качество производительности измеряется в соответствии с показателями эффективности по ГОСТ Р ИСО/МЭК 25022—2021 (подраздел 8.3).

16 Удовлетворенность

16.1 Удовлетворенность

Качество удовлетворенности измеряется в соответствии с показателями общей удовлетворенности по ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.4.1).

16.2 Полноценность

Качество полноценности измеряется в соответствии с показателями полезности по ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.4.2).

16.3 Доверие

Качество доверия измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.4.3).

16.4 Удовольствие

Качество удовольствия измеряется в соответствии с показателями удовольствия пользователя по ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.4.4).

16.5 Комфорт

Качество комфорта измеряется в соответствии с показателями эргономического комфорта по ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.4.5).

16.6 Прозрачность

Рекомендуемое руководство по прозрачности приведено в 9.8.

17 Свобода от риска

Риски, возникающие при использовании ИИ, должны быть идентифицированы, описаны в качественном и количественном отношении, их ранжируют в соответствии с критериями риска и целями организации, в рамках управления рисками систем ИИ.

Риски, возникающие при разработке, использовании систем ИИ и управлении ими, приведены в [8].

17.1 Смягчение отрицательных последствий экономического риска

Качество смягчения отрицательных последствий экономического риска измеряется в соответствии с показателями снижения экономического риска по *ГОСТ Р ИСО/МЭК 25022—2021* (пункт 8.5.2).

17.2 Смягчение отрицательных последствий риска здоровья и безопасности

Качество смягчения отрицательных последствий риска здоровья и безопасности измеряется в соответствии с показателями снижения рисков здоровья и безопасности по *ГОСТ Р ИСО/МЭК 25022—2021* (пункт 8.5.3).

Достижение снижения рисков для здоровья и безопасности должно быть постоянной деятельностью на протяжении всего жизненного цикла продукта системы, включая раннюю концепцию, разработку требований, проектирование, разработку, управление качеством, поставку, приобретение, техническое обслуживание и вывод из эксплуатации (см. [2]). Фаза требований особенно важна при разработке более безопасных систем ИИ.

Другие стандарты и методы по-разному касаются здоровья и безопасности и могут дополнять требования и рекомендации, представленные в настоящем стандарте.

Устранение рисков для здоровья и безопасности требует уравнивания часто конкурирующих соображений в контексте, в том числе в отношении других характеристик качества, таких как безопасность.

Несколько других характеристик качества, подробно описанных в настоящем стандарте, могут иметь значительное положительное или отрицательное влияние на безопасность системы ИИ — к ним относятся удобство использования (раздел 9), обучаемость (раздел 9.2), работоспособность (раздел 9.3), надежность (раздел 10) и функциональная пригодность (раздел 6).

Следовательно, разработчики и пользователи систем ИИ должны тщательно взвесить относительную важность выбранных ими характеристик качества.

Для эффективного снижения рисков безопасности организация должна:

- определить требования по охране здоровья и безопасности своих систем в соответствии с предполагаемым использованием;
- применять методы снижения риска для здоровья и безопасности на системном уровне;
- определить требования к качеству программного обеспечения, вытекающие из системных требований к системе ИИ для здоровья и безопасности на системном уровне и прослеживаемые к ним;
- выявлять и устранять риски для здоровья и безопасности, связанные с человеческими, процессными и техническими элементами их систем;
- внедрить и эксплуатировать эффективную организационную систему управления безопасностью, охватывающую весь жизненный цикл системы ИИ;
- интегрировать свою систему управления программным обеспечением в систему управления качеством ИИ организации (см. [9]);
- заявить о намерении улучшить меры по снижению рисков для здоровья и безопасности;
- обеспечить, чтобы деятельность по разработке предметной области и требований выполнялась достаточно междисциплинарной группой экспертов;
- идентифицировать:
 - а) среду, в которой предполагается использовать их системы;
 - б) цели его системы ИИ;
 - в) ограничения безопасности (поведенческие границы), в пределах которых его системы ИИ могут безопасно работать;
- внедрить меры контроля, чтобы гарантировать, что ограничения по охране труда и технике безопасности не нарушаются;
- реализовать меры:
 - а) по выявлению и предотвращению нежелательных потерь;
 - б) устранению и снижению опасностей, в том числе за счет проектирования системы, языка программирования и выбора метода разработки;
- постоянно узнавать о снижении рисков для здоровья и безопасности;
- выявлять опережающие и отстающие показатели здоровья и безопасности и реагировать на них;
- провести тщательную оценку качества анализа требований к системе ИИ, чтобы убедиться, что он включает:
 - а) определение целей в области охраны здоровья и безопасности;

- б) определение ограничений безопасности;
- в) выявление режимов отказов взаимодействия компонентов (включая режимы отказов отдельных компонентов) в виде социальных, процедурных и технических форм;
 - осуществлять проектирование человеческого фактора при анализе требований к системам, а также при проектировании и эксплуатации систем;
 - перечислить подробные предположения об указанной предполагаемой области и предполагаемом рабочем использовании системы ИИ.

Примечание — Организации следует учитывать расхождение перечисленных допущений с областями и областью применения, в которой она планирует использовать системы ИИ, в качестве источника риска, который необходимо устранить, и в качестве фактора выбора системы;

- провести анализ опасностей, включая методы, которые определяют различные формы предубеждений, в том числе когнитивные предубеждения у предполагаемых пользователей системы;
- выявление и устранение уязвимостей, связанных с неотъемлемыми предубеждениями в используемых методах анализа опасностей (см. ПНСТ 839—2023);
- спроектировать системы ИИ для работы в безопасных состояниях и, по возможности, восстанавливаться из небезопасных состояний.

Организация должна предотвращать следующие ситуации:

- отсутствие значимого надзора за выбором, проектированием, разработкой, эксплуатацией, обслуживанием или выводом из эксплуатации системы ИИ (см. ПНСТ 843—2023);
- отсутствие конкретных целей и требований безопасности;
- требования безопасности, разработанные после проектирования системы ИИ;
- задокументированные предположения, которые не соответствуют предполагаемому использованию системы ИИ;
- отсутствие задокументированных допущений;
- объединение других характеристик качества с характеристиками безопасности, например:
 - а) резервирование;
 - б) надежность;
 - в) удобство использования.

Примечание — Как указано во введении к настоящему пункту, эти (и другие) характеристики качества могут иметь существенное влияние на здоровье и безопасность, связанные с системой ИИ, но сами по себе они не составляют безопасность. То есть надежная система ИИ не обязательно является безопасной системой (действительно, в некоторых случаях менее надежная система может быть безопаснее надежной);

- опасности, которые были идентифицированы, но затем отброшены как «маловероятные», а не спланированы или, если это невозможно, смягчены в максимально возможной степени;
- принятие урезанных методов анализа опасности для приемлемости риска системы ИИ.

Примечание — Опасности должны быть проанализированы;

- использование параметров или методов, которые не могут быть проверены (например, расчеты вероятности) для принятия решений о рисках для безопасности полетов (т. е. о том, какие риски следует учитывать);
- сосредоточить внимание на выявлении и снижении рисков, а не на выявлении и устранении опасностей;
- фокус/идентификация ошибки пользователя/оператора;
- утверждения/заявления о том, что система ИИ или ее компоненты являются «безопасными».

Примечание — Для целей настоящего стандарта безопасность рассматривается как эмерджентное свойство социотехнической системы (с человеческими, процедурными и техническими компонентами), возникающее в результате взаимодействия ее составных частей. Таким образом, невозможно оценить систему как безопасную или небезопасную;

- использование формулировок в документации, указывающих на принятие моделей потерь (несчастных случаев) с линейной причинно-следственной связью, таких как «цепочка событий» и «первопричина».

Примечание — Важно, чтобы надзора было достаточно для выявления и устранения сбоев в системе.

17.3 Смягчение отрицательных последствий экологического риска

Качество смягчения отрицательных последствий экологического риска измеряется в соответствии с показателями снижения экологических рисков по *ГОСТ Р ИСО/МЭК 25022—2021* (пункт 8.5.4).

Организация должна идентифицировать, в свете влияния на экологию, потенциальные прямые и косвенные преимущества и недостатки, связанные с использованием систем ИИ (например, потребление энергии, выбросы парниковых газов, потребление воды). В дополнение к типичным экологическим рискам, рассматриваемым для любой компьютерной системы, разработчики системы ИИ должны принимать во внимание следующее:

- вычислительные ресурсы, базы данных и приложения, используемые при разработке и функционировании системы ИИ;
- детализацию функциональных требований и задач функциональной корректности, поскольку они могут потребовать использования системы ИИ большой мощности, что приведет к более значительному негативному воздействию на окружающую среду. Следует оценить уровень компьютерных ресурсов, используемых для обучения системы, получения логического вывода и для ее функциональной адаптируемости, поскольку чем больше ресурсов задействовано и чем дольше они используются, тем более негативным будет воздействие системы ИИ на окружающую среду;
- выбор стратегии адаптируемости системы ИИ, поскольку он может приводить к дополнительным рискам в сфере экологии. Например, система, в которой адаптация происходит непрерывно, может потреблять больше энергии, чем та, в которой адаптация происходит менее часто. Точно так же использование системы ИИ, которая настраивается путем непрерывного обучения, более эффективно, чем проведение обучения заново полностью.

План оценки рисков и их предотвращения должен также включать косвенные риски:

- ресурсы, используемые на изготовление аппаратных средств для разработки системы ИИ и управление ею (включая использование аппаратных средств общего назначения в противоположность средствам, разработанным для конкретных типов систем ИИ). Организация должна принимать во внимание риски, относящиеся ко всей цепи ее поставщиков;
- воздействие систем ИИ, поскольку применение ИИ и его структурное и поведенческое влияние может приводить как к положительному, так и к отрицательному воздействию на окружающую среду. Например, система ИИ, которая оптимизирует процесс добычи полезных ископаемых, может усугубить негативное воздействие добывающей, перерабатывающей и промышленной отраслей. Система ИИ может рекомендовать покупку какого-либо продукта в интернет-торговле, что может привести к росту неумеренного потребления. Но система ИИ, которая оптимизирует маршруты движения транспорта, оказывает положительное влияние на окружающую среду за счет снижения расхода энергии, или система, которая прогнозирует спрос на скоропортящиеся товары, может повлиять на снижение объема отходов.

Выявленные риски должны изучаться, так, чтобы меры по их снижению стали частью требований к системе.

17.4 Снижение отрицательных последствий социального и этического риска

Организация, которая разрабатывает, производит, развертывает или использует системы ИИ, должна управлять социальными и этическими рисками (см. [8], в котором содержатся рекомендации по ключевым общественным или этическим аспектам).

Организация должна выявлять и решать общественные и этические проблемы на протяжении всего жизненного цикла систем ИИ в рамках процесса снижения рисков. Дополнительную информацию см. в *ПНСТ 839—2023*, *ПНСТ 840—2023* и [8].

18 Покрытие контекста

18.1 Полнота контекста

Качество полноты контекста измеряется в соответствии с показателями, приведенными в *ГОСТ Р ИСО/МЭК 25022—2021* (пункт 8.6.2).

Одной из главных отличительных черт систем ИИ является их слабая способность объяснить происходящие внутри процессы пользователю (объясняемость) и продемонстрировать связь параметров с полученным результатом (интерпретируемость), особенно это справедливо в отношении моделей

МО. Это подразумевает более детальное описание контекста использования для системы ИИ, чем для обычной системы. Описание контекста использования может быть представлено как набор значимых факторов жизненного цикла с детальными предварительными расчетами статистических характеристик (функции плотности вероятности) для каждого фактора на протяжении всего жизненного цикла системы ИИ.

При оценке качества системы ИИ следует исходить из значимых факторов жизненного цикла для одной ситуации или сочетания обстоятельств, возможностей или индивидуальных предпочтений.

Такие детальные предварительные расчеты могут быть получены исходя из предполагаемых условий эксплуатации, технических спецификаций, жизненного опыта и т. п. Для разных значимых факторов могут использоваться разные типы шкалы измерений (числовая шкала, интервальная шкала, категориальная шкала, шкала отношений и т. д.). При расчете статистических характеристик значимых факторов должны приниматься во внимание особенности кросс-корреляции этих факторов.

Например, для системы ИИ, распознающей лица, значимые факторы жизненного цикла включают:

- перспективное изображение ракурсов лица (три ракурса);
- пол, возраст и расовую принадлежность человека;
- наличие зрительных помех и относительную площадь части лица, закрытой предметом (очками, головным убором, бородой и т. п.);
- временной интервал между стадиями обучения и выполнением реальной задачи моделью МО;
- уровень и градиент освещенности лица;
- характеристики используемого сенсора.

Полнота списка значимых факторов жизненного цикла и точность предварительных расчетов статистических характеристик этих факторов определяют качество описания контекста использования.

Описание контекста использования может использоваться для генерации тестовых наборов данных для оценки характеристик системы ИИ. В этом случае качество описания определяет репрезентативность измеряемых характеристик системы ИИ.

Пусть $\bar{F} = \{f_i\}$ — это набор значимых факторов жизненного цикла $f_i, i = 1..N$ для конкретной системы ИИ; $D(\bar{F})$ — это функция плотности вероятности от $\bar{F}$.

Если значимые факторы жизненного цикла статистически независимы, тогда функцию $D(\bar{F})$ можно представить как $D(\bar{F}) = \prod_i^N d_i(f_i)$, где $d_i(f_i)$ — это функция плотности вероятности от f_i .

Пусть конкретные контексты использования (конкретные условия) для системы ИИ определяются как диапазон значений

$$\Omega_1 \in \bar{F} \quad (1)$$

и контексты за пределами конкретных условий использования определяются как диапазон значений

$$\Omega_0 \in \bar{F} \quad (2)$$

таким образом, что

$$\Omega_1 \cap \Omega_0 = \emptyset, \quad (3)$$

$$\Omega_1 \cup \Omega_0 = \Omega, \quad (4)$$

где Ω — диапазон всех возможных значений значимых факторов жизненного цикла $\bar{F}$.

Пусть $\bar{F}$ принимает дискретные значения так, что

$$L_1 = |\Omega_1|, \quad (5)$$

$$L_0 = |\Omega_0|, \quad (6)$$

где L_1 и L_0 — мощность множеств Ω_1 и Ω_0 соответственно.

Пусть функция качества системы ИИ $Q(\bar{F})$ определена так, что

$$Q(\bar{F}) = \begin{cases} 1, & \text{если качество системы ИИ удовлетворяет требованиям,} \\ & \text{заданным пользователем,} \\ 0, & \text{если качество системы ИИ не удовлетворяет заданным} \\ & \text{требованиям.} \end{cases} \quad (7)$$

Тогда покрытие контекста может быть рассчитано как среднее значение величины показателя качества во всех контекстах в соответствии с уравнением

$$C = \frac{\sum_{\bar{F} \in \Omega}^L Q(\bar{F}) D(\bar{F})}{\sum_{\bar{F} \in \Omega}^L D(\bar{F})} = \sum_{\bar{F} \in \Omega}^L Q(\bar{F}) D(\bar{F}), \quad (8)$$

где $L = L_1 + L_0$ — мощность множества Ω .

C изменяется в диапазоне от 0 до 1, где 0 соответствует непригодной для эксплуатации системе ИИ, а 1 соответствует системе ИИ надлежащего качества во всех вероятных контекстах использования.

Гибкость может быть рассчитана как среднее значение величины показателя качества за пределами заданного контекста в соответствии с уравнением

$$C_0 = \frac{\sum_{\bar{F} \in \Omega_0}^{L_0} Q(\bar{F}) D(\bar{F})}{\sum_{\bar{F} \in \Omega_0}^{L_0} D(\bar{F})} = \frac{\sum_{\bar{F} \in \Omega_0}^{L_0} Q(\bar{F}) D(\bar{F})}{1 - D_1}. \quad (9)$$

Покрытие контекста, полнота контекста и гибкость объединяются следующим уравнением:

$$C = D_1 C_1 + (1 - D_1) C_0. \quad (10)$$

Обычно для системы ИИ, хорошо адаптированной к заданному контексту использования

$$C_0 < C < C_1. \quad (11)$$

18.2 Гибкость

Качество гибкости измеряется в соответствии с показателями, приведенными в ГОСТ Р ИСО/МЭК 25022—2021 (пункт 8.6.3).

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 ИСО/МЭК 25022—2021	IDT	ISO/IEC 25022:2016 «Системная и программная инженерия. Требования и оценка качества систем и программных средств (SQuaRE). Измерение качества при использовании»
ГОСТ Р ИСО/МЭК 25023—2021	IDT	ISO/IEC 25023:2016 «Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Измерение качества системы и программной продукции»
ГОСТ Р ИСО/МЭК 25040—2014	IDT	ISO/IEC 25040:2011 «Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Процесс оценки»
ГОСТ Р МЭК 61508-2—2012	IDT	IEC 61508-2:2010 «Системы электрические/электронные/программируемые электронные, связанные с функциональной безопасностью. Часть 2. Требования к электрическим/электронным/программируемым электронным системам, связанным с безопасностью»
ГОСТ Р 70462.1— 2022/ ISO/IEC TR 24029-1—2021	IDT	ISO/IEC TR 24029-1:2021 «Искусственный интеллект (ИИ). Оценка робастности нейронных сетей. Часть 1. Обзор»
ПНСТ 835—2023 (ISO/IEC TS 4213:2022)	MOD	ISO/IEC TS 4213:2022 «Информационные технологии. Искусственный интеллект. Оценка эффективности классификации машинного обучения»
ПНСТ 837—2023/ ISO/IEC TS 8200	IDT	ISO/IEC TS 8200 «Искусственный интеллект. Управляемость автоматизированных систем искусственного интеллекта»
ПНСТ 838—2023 (ISO/МЭК 23053:2022)	MOD	ISO/IEC 23053:2022 «Платформа разработки систем искусственного интеллекта (AI) с использованием машинного обучения (ML)»
ПНСТ 839—2023 (ISO/IEC TR 24027:2021)	MOD	ISO/IEC TR 24027:2021 «Информационные технологии. Искусственный интеллект (ИИ). Смещенность в системах ИИ и при принятии решений с помощью ИИ»
ПНСТ 840—2023 (ISO/IEC TR 24368:2022)	MOD	ISO/IEC TR 24368:2022 «Информационная технология. Искусственный интеллект. Обзор этических и социальных проблем»
ПНСТ 842—2023 (ISO/МЭК 25059:2023)	MOD	ISO/IEC 25059:2023 «Программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модель качества для систем на основе искусственного интеллекта»
ПНСТ 843—2023 (ISO/МЭК 38507:2022)	MOD	ISO/IEC 38507:2022 «Информационная технология. Управление ИТ. Последствия использования искусственного интеллекта организациями для управления информационными технологиями»
Примечание — В настоящей таблице использованы следующие условные обозначения степени соответствия стандартов: - IDT — идентичные стандарты; - MOD — модифицированные стандарты.		

Библиография

- [1] Серия ISO/IEC/IEEE 29119 Системная и программная инженерия. Тестирование программного обеспечения (Software and systems engineering — Software testing)
- [2] ИСО/МЭК 22989:2022 Информационная технология. Искусственный интеллект. Концепции и терминология искусственного интеллекта (Information technology — Artificial intelligence — Artificial intelligence concepts and terminology)
- [3] ISO/IEC/IEEE 29119-1:2022 Системная и программная инженерия. Тестирование программного обеспечения. Часть 1. Основные понятия (Software and systems engineering — Software testing — Part 1: General concepts)
- [4] ISO/IEC/IEEE 29148:2018 Программная и системная инженерия. Процессы жизненного цикла. Разработка требований (Systems and software engineering — Life cycle processes — Requirements engineering)
- [5] ИСО/МЭК 24029-2:2023 Искусственный интеллект (ИИ). Оценка робастности нейронных сетей. Часть 2. Методология использования формальных методов (Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods)
- [6] ИСО 21448:2022 Дорожные транспортные средства. Безопасность заданных функций (Road vehicles — Safety of the intended functionality)
- [7] ИСО 26262-11 Дорожные транспортные средства. Функциональная безопасность. Часть 11. Руководящие указания по применению ISO 26262 к полупроводникам (Road vehicles — Functional safety — Part 11: Guidelines on application of ISO 26262 to semiconductors)
- [8] ИСО/МЭК 23894:2022 Информационная технология. Искусственный интеллект. Руководство по менеджменту риска (Information technology — Artificial intelligence — Guidance on risk management)
- [9] ИСО/МЭК 42001 Искусственный интеллект. Система управления (Information technology — Artificial intelligence — Management system)

УДК 004.01:006.354

ОКС 35.030

Ключевые слова: системная и программная инженерия, качество систем и программного обеспечения, оценка качества, системы искусственного интеллекта

Редактор *Е.Ю. Митрофанова*
 Технический редактор *И.Е. Черепкова*
 Корректор *Л.С. Лысенко*
 Компьютерная верстка *И.Ю. Литовкиной*

Сдано в набор 17.11.2023. Подписано в печать 30.11.2023. Формат 60×84%. Гарнитура Ариал.
 Усл. печ. л. 2,79. Уч-изд. л. 2,51.

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