**ФЕДЕРАЛЬНОЕ АГЕНТСТВО**

**ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ**

|  |  |  |
| --- | --- | --- |
| Znak_nacionalnoi_standartizacii | **НАЦИОНАЛЬНЫЙ**  **СТАНДАРТ**  **РОССИЙСКОЙ**  **ФЕДЕРАЦИИ** | **ГОСТ Р**  *(проект,*  *окончательная*  *редакция)* |

ОЦЕНКА КАЧЕСТВА

СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

**Общие положения**

**Издание официальное**

*Настоящий проект стандарта не подлежит применению до его утверждения*

**Москва**

**Стандартинформ**

**2021**

# Предисловие

1 РАЗРАБОТАН Национальным исследовательским университетом «Высшая школа экономики» в соответствии с Программой национальной стандартизации на 2020 г.

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 ….

4 ВВЕДЕН ВПЕРВЫЕ

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

© Стандартинформ, оформление, 20\_\_

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

# Содержание

Предисловие

Содержание

Введение

1 Область применения

2 Нормативные ссылки

3 Термины

4 Сокращения

5 Модель качества СИИ .

6 Методология и критерии оценки качества на стадиях жизненного цикла СИИ

7 Общие принципы и порядок оценки качества СИИ

7.1. Разработка программы тестирования (испытаний)

7.1.1 Общие требования к программе тестирования (испытаний)

7.1.2 Создание экспертной группы

7.1.3 Определение требований к качеству

7.1.4 Выбор и обоснованность существенных характеристик (субхарктеристик)

7.2 Алгоритм оценки качества СИИ

7.2.1 Подготовительные работы

7.2.2 Тестирование и оценка показателей качества

7.2.3 Анализ и интерпретация результатов тестирования

8 Представительный набор существенных характеристик и показатели качества СИИ

8.1 Общие требования и критерии выбора показателей (метрик) качества СИИ

8.2 Функциональная пригодность (functionality). Метрики

8.3 Функциональная производительность (performance efficiency). Метрики

8.4 Совместимость (сompatibility). Метрики

8.5 Мобильность (portability). Метрики

8.6 Практичность (usability). Метрики

8.7 Сопровождаемость (maintainability). Метрики

8.8 Надежность (reliability). Метрики

9 Набор входных и выходных данных

9.1 Требования к набору входных и выходных данных для тестирования СИИ

**Введение**

Оценка качества является неотъемлемой частью жизненного цикла систем искусственного интеллекта (СИИ) и включает в себя действия, проводимые на этапах разработки, выпуска и эксплуатации данных систем в целях обеспечения необходимого уровня соответствия СИИ предъявляемым требованиям. В зависимости от этапа жизненного цикла оценка качества позволяет:

- определить текущие параметры СИИ и выполнить действия, направленные на повышение ее надежности и производительности, а также расширения функциональности системы (путем выполнения процедур по устранению неисправностей и совершенствованию программного обеспечения);

- удостовериться, что выходные данные СИИ являются приемлемыми и обоснованными для решения поставленной задачи в условиях, представленных в описании СИИ, с учетом современного уровня развития отрасли;

- убедиться в достижении целей предназначения СИИ в условиях обеспечения заданной точности, надежности и достоверности выходных данных;

- подтвердить соответствие характеристик СИИ требуемым значениям, установленным в технической документации и/или нормативных правовых актах.

Отсутствие надлежащей оценки качества систем ИИ может привести к угрозам безопасности людей, окружающей природной среды, материальных и нематериальных активов. С другой стороны, наличие грамотно выстроенной системы оценки качества позволяет повысить доверие к системам ИИ на физическом уровне путем подтверждения требований к надежности, безопасности и функциональности.

Настоящий стандарт разработан на основе ИСО/МЭК 25010 «Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов», определяющего общие подходы к оценке качества программных продуктов и преимущественно программных вычислительных систем, но не учитывающего специфику вычислительных алгоритмов и характеристик СИИ.

Настоящий стандарт позволяет дополнить представленный в ИСО/МЭК 25010 набор показателей качества специализированными характеристиками (субхарактеристиками) и соответствующими метриками для обеспечения полноценной оценки качества СИИ.

В настоящем стандарте: определено понятие качества СИИ; приведена методология, показатели и критерий оценки качества на стадиях жизненного цикла СИИ;

формализована модель качестваСИИ; приведена классификация существенных характеристик и показателей качества СИИ.

**НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Оценка качества**

**систем искусственного интеллекта**

**Общие положения**

Artificial intelligence systems. Quality assurance. General

**Дата введения — 20 – –**

## 1 Область применения

Настоящий стандарт устанавливает общие требования к оценке качества СИИ, включая:

- виды существенных характеристик СИИ, подтверждение значений которых установленным требованиям обеспечивает доверие к этим системам;

- порядок выбора и оценки метрик качества на основании представленного набора существенных характеристик;

- требования к набору входных и выходных данных для тестирования СИИ;

- вопросы планирования, разработки программы и процедуры оценки качества СИИ, позволяющих провести надлежащую валидацию и верификацию продукта в целях подтверждения его надежности, безопасности и функциональности.

Принципы, установленные в стандарте, применимы к СИИ в различных отраслях и их следует придерживаться, насколько это возможно, с учетом отраслевых и национальных регулирующих требований.

Требования к наборам данных и показателям качества систем СИИ, приведенные в настоящем стандарте, могут быть дополнены требованиями конкретных стандартов в соответствующих областях применения СИИ (например, ГОСТ Р 52633.1-2010 и ГОСТ Р 52633.2-2010).

Настоящий стандарт распространяется на все системы, использующие различные методы искусственного интеллекта (ИИ), включая алгоритмы на основе машинного обучения (обучение по прецедентам) и экспертные системы (на основе дедуктивного обучения).

|  |
| --- |
| **Издание официальное** |

## 2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р ИСО 9000-2015 Системы менеджмента качества. Основные положения и словарь (Издание с Поправкой);

ГОСТ Р 54837-2011 Информационная технология. Обучение, образование и подготовка. Менеджмент качества, обеспечение качества и метрики. Часть 3. Эталонные методы и метрики;

ГОСТ Р ИСО/МЭК 9126-93 Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению;

ГОСТ Р ИСО/МЭК 25010-2015 Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов;

ГОСТ Р ИСО 5725-1-2002 Точность (правильность и прецизионность) методов и результатов измерений. Часть 1. Основные положения и определения;

ГОСТ Р 52633.1-2010 Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации;

ГОСТ Р 52633.2-2010 Защита информации. Техника защиты информации. Требования к формирована синтетических биометрические образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации.

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 следующие термины с соответствующими определениями.

**3.1 Общие определения**

3.1.1 **Аннотирование данных** (data annotation): Процесс маркирования данных, выполняемый для того, чтобы сделать данные пригодными для машинного обучения.

3.1.2 **Базовое значение показателя качества:** значение показателя качества системы искусственного интеллекта, принятое за основу при сравнительной оценке ее качества.

3.1.3 **Безопасность:** свойство системы искусственного интеллекта сохранять состояние, характеризующееся отсутствием недопустимого риска, при использовании ее по назначению в условиях, предусмотренных изготовителем.

3.1.4

|  |
| --- |
| **Валидация (validation)**: Подтверждение посредством представления объективных свидетельств того, что требования, предназначенные для конкретного использования или применения, выполнены.  Примечание — Валидация в контексте жизненного цикла представляет собой совокупность действий, гарантирующих и обеспечивающих уверенность в том, что система способна реализовать свое предназначение, текущие и перспективные цели.  [ГОСТ Р ИСО/МЭК 12207-2010, пункт 4.54] |

3.1.5 **Выборка (sample):** набор данных, представляющий собой подмножество генеральной совокупности.

3.1.6.

|  |
| --- |
| **Выброс (outlier):** Элемент маломощного подмножества выборки, существенно отличающийся от остальных элементов выборки.  Примечание 1 - Классификация наблюдения или подмножество выборки как выброс (или выбросы) зависит от выбранной модели генеральной совокупности, из которой отобрана выборка. Выброс не рассматривают как истинный элемент генеральной совокупности.  Примечание 2 - Выброс может появиться из другой генеральной совокупности, быть результатом некорректной регистрации данных или общей ошибкой измерений. Примечание 3 - Подмножество может содержать одно или несколько наблюдений.  [ГОСТ Р ИСО 16269-4-2017, пункт 2.1] |

3.1.7

|  |
| --- |
| **Верификация (verification)**: Подтверждение посредством представления объективных свидетельств того, что установленные требования были выполнены.  Примечание — Верификация в контексте жизненного цикла представляет собой совокупность действий по сравнению полученного результата жизненного цикла с требуемыми характеристиками для этого результата. Результатами жизненного цикла могут являться (но не ограничиваться ими): заданные требования, описание проекта и непосредственно система.  [ГОСТ Р ИСО/МЭК 12207-2010, пункт 4.55] |

3.1.8 **Генеральная совокупность:** Множество всех рассматриваемых прецедентов.

3.1.9

|  |
| --- |
| **Жизненный цикл (life cycle)**: Развитие системы, продукции, услуги, проекта или другой создаваемой изготовителем сущности от замысла до списания.  [ГОСТ Р 57193-2016, пункт 4.1.19, с изменениями] |

3.1.10

|  |
| --- |
| **Искусственный интеллект, ИИ**: Способность технической системы имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.  [ГОСТ Р 59276-2020, пункт 3.6] |

3.1.11

|  |
| --- |
| **Изготовитель (производитель):** Физическое или юридическое лицо, несущее ответственность за разработку, производство, упаковку и маркировку изделия, прежде чем оно займет место на рынке под собственным наименованием независимо от того, выполняются эти действия данным лицом непосредственно или привлеченным третьим лицом.  [ГОСТ Р 57501-2017, пункт 3.18] |

3.1.12 **Качество:** совокупность характеристик и свойств СИИ, обусловливающих её способность удовлетворять установленным или предполагаемым требованиям в соответствии с ее назначением.

3.1.13 **Критерий оценки качества (quality assessment criterion):** Набор определенных и задокументированных правил и условий, которые используются для решения о приемлемости общего качества конкретной системы искусственного интеллекта.

3.1.14

|  |
| --- |
| **Метаданные (metadata)**: Информация о ресурсе.  [ГОСТ 57668-2017 (ИСО 19115-1:2014), пункт 4.10] |

3.1.15 **Метрика (metric):** материальная мера некоторых аспектов характеристик качества.

3.1.16 **Набор данных**: Совокупность данных, в том числе соответствующих им метаданных, организованных по определенным правилам и принципам описания.

Примечание – В зависимости от цели применения набор данных может быть представлен следующими типами данных: текстовыми записями, временными рядами, изображениями, видео, сигналами и т.п.

3.1.17 **Надежность:** свойство системы искусственного интеллекта сохранять во времени способность выполнять требуемые функции при использовании ее по назначению в условиях, предусмотренных изготовителем.

3.1.18 **Показатель качества системы искусственного интеллекта**: Степень соответствия представительного набора существенных (значимых) характеристик системы искусственного интеллекта требованиям, то есть потребностям или ожиданиям, которые установлены, обычно предполагаются или являются обязательными для этой системы.

3.1.19 **Оптимальное значение показателя качества**: Значение показателя качества системы искусственного интеллекта, при котором достигается либо наибольший эффект от ее применения при заданных затратах на ее создание и эксплуатацию, либо заданный эффект при наименьших затратах, либо наибольшее отношение эффекта к затратам.

3.1.20 **Оценка качества:** совокупность операций, включающих выбор номенклатуры показателей качества, определение значений этих показателей и сравнение их с базовыми значениями

3.1.21 **Предвзятость, необъективность (bias):** Свойство системы искусственного интеллекта, заключающееся в принятии ошибочных решений, связанных со статистической смещённостью обучающей выборки исходных данных.

3.1.22 **Представительный набор существенных характеристик:** Минимально необходимая и достаточная совокупность характеристик системы искусственного интеллекта, позволяющая потребителю, организациям, ответственным за регулирование вопросов создания и применения систем искусственного интеллекта, или любой другой заинтересованной стороне достоверно оценивать качество системы при решении конкретной прикладной задачи.

3.1.23 **Прецедент:** множество элементов (объектов, событий, последовательности действий и т. п.), используемых для описания поведения системы.

3.1.24

|  |
| --- |
| **Ресурс (resource):** Идентифицируемый аспект или средства, которые соответствуют требованиям.  Пример - Набор данных, комплект наборов данных, сервис, документ, инициатива, программное обеспечение, лицо или организация.  [ГОСТ 57668-2017 (ИСО 19115-1:2014), пункт 4.17] |

3.1.25

|  |
| --- |
| **Сервис (service):** Отдельная часть функциональности, предоставляемой сущностью посредством интерфейса.  [ГОСТ 57668-2017 (ИСО 19115-1:2014), пункт 4.18] |

3.1.26

|  |
| --- |
| **Система искусственного интеллекта (artificial intelligence system):** Техническая система, в которой используются технологии ИИ и обладающая ИИ.  [ГОСТ Р 59276-2020, пункт 3.16] |

3.1.27 **Существенные (значимые) характеристики системы искусственного интеллекта (СИИ)**: Характеристики системы искусственного интеллекта, определяющие его функциональность, надежность и безопасность при решении конкретной прикладной задачи, подтверждение соответствия которых установленным требованиям может быть выполнено потребителем системы, организациями, ответственными за регулирование вопросов создания и применения систем искусственного интеллекта, или любой другой заинтересованной стороной. Характеристики систем искусственного интеллекта, подтверждение соответствия которых установленным требованиям может быть выполнено исключительно разработчиком системы, не относятся к существенным;

3.1.28 **Техническая система:** целостная совокупность конечного числа взаимосвязанных материальных объектов, имеющая последовательно взаимодействующие сенсорную и исполнительную функциональные части, модель их предопределенного поведения в пространстве равновесных устойчивых состояний и способная при нахождении хотя бы в одном из них (целевом состоянии) самостоятельно в штатных условиях выполнять предусмотренные ее конструкцией потребительские функции [10].

3.1.29

|  |
| --- |
| **Технологии искусственного интеллекта (ИИ)**: комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека при решении задач компьютерного зрения, обработки естественного языка, распознавания и синтеза речи, поддержки принятия решений и других практически значимых задач обработки данных.  [ГОСТ Р 59276-2020, пункт 3.20] |

3.1.30

|  |
| --- |
| **Точность (accuracy):** Степень близости результата испытаний к принятому опорному значению.  [ГОСТ Р ИСО 5725-1-2002, пункт 3.6] |

## 4 Сокращения

В настоящем стандарте применены следующие обозначения и сокращения:

ИИ — искусственный интеллект;

ЖЦ – жизненный цикл;

СИИ — система искусственного интеллекта;

ЦПУ – центральное процессорное устройство.

## 5 Модель качества

Тестирование СИИ в процессе ее жизненного цикла осуществляется с целью оценки соответствия системы требованиям функциональной, надежности и безопасности СИИ.

Качество определяет способность СИИ при заданных условиях удовлетворять установленным или предполагаемым потребностям различных заинтересованных сторон, что позволяет, таким образом, оценить его достоинства.

**5.1 Общие положения**

5.1.1 Для описания качества СИИ используется модель качества, представляющая собой структурированное множество характеристик, субхарактеристик, метрик и отношений между ними.

5.1.2 Для оценки качества СИИ используются следующие модели качества по ГОСТ Р ИСО/МЭК 25010:

- модель качества при использовании;

- модель качества продукта;

- модель качества данных.

Примечание– В настоящем стандарте подробно рассмотрена модель качества СИИ продукта.

5.1.3 Качество СИИ, как любого программного обеспечения, является интегральным показателем, для оценки которого на этапе построения модели качества (рис. 1), исходя из функционального назначения системы и решаемой прикладной интеллектуальной задачи, определяется представительный набор показателей качества (существенных характеристик и субхарактеристик).

где S – оцениваемая СИИ,

F – представительный набор существенных характеристик,

fi – i-ая существенная характеристика.

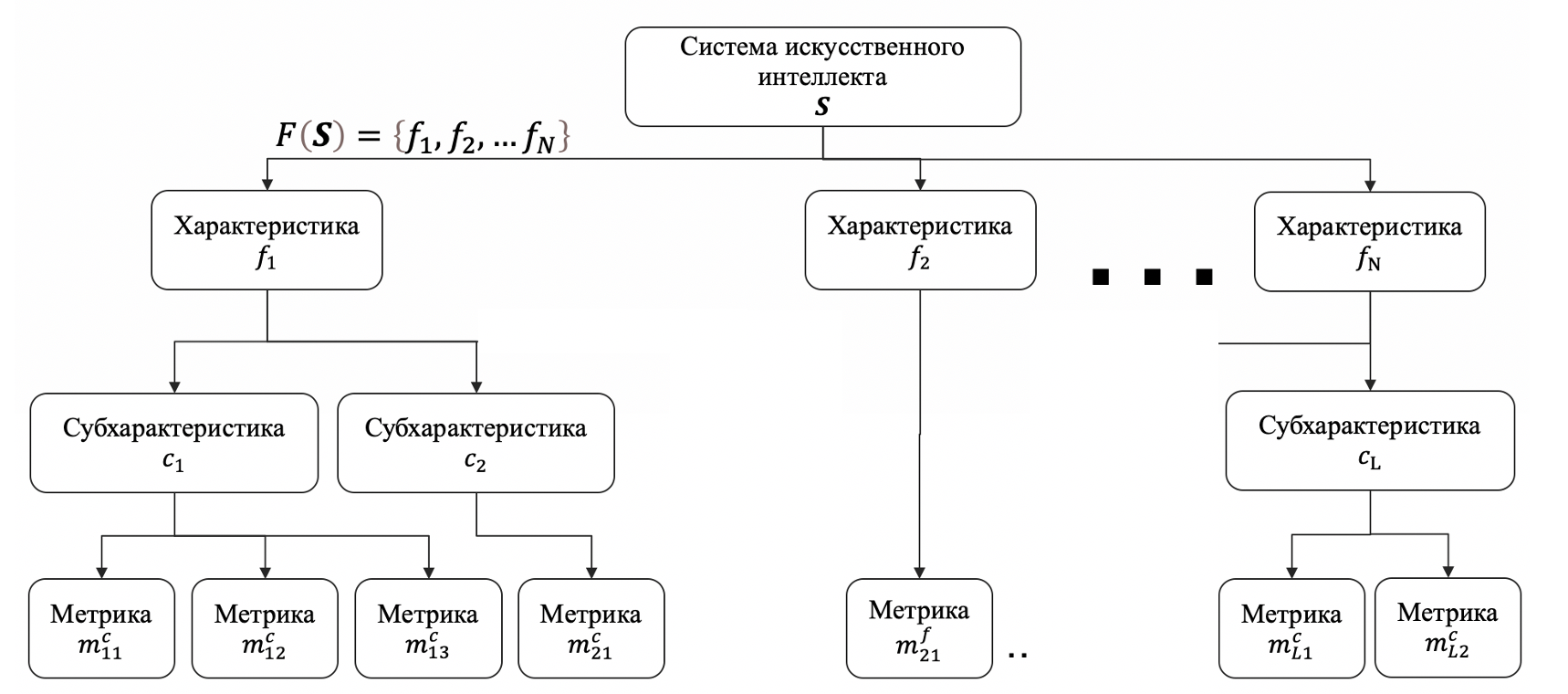


Рисунок 1 – Дерево существенных характеристик и метрик СИИ

5.1.4 Для количественного измерения показателя качества (характеристики fi, субхарактеристики сi) используется соответствующая метрика mi, определяемая на измерительной шкале, тип которой выбирается исходя из физического смысла соответствующей характеристики fi (числовая шкала, шкала категорий, отношений и др.).

**6 Методология и критерии оценки качества на стадиях жизненного цикла СИИ**

**6.1 Общие положения**

6.1.1 Оценка качества осуществляется в течение всего жизненного цикла СИИ при:

* планировании показателей качества СИИ;
* контроле качества на отдельных этапах разработки СИИ по ГОСТ 34.601-90 (техническое задание, технический проект, рабочая документация);
* контроле качества в процессе производства СИИ;
* проверке эффективности модификации СИИ на этапе сопровождения.

6.1.2 В зависимости от этапа жизненного цикла СИИ тестирование могут выполнять различные группы лиц.

Таблица 1 – Специалисты, осуществляющие тестирование СИИ

|  |  |  |
| --- | --- | --- |
| № п/п | Этап жизненного цикла | Специалисты,  осуществляющие тестирование СИИ |
| 1 | Разработка (производство) | Производитель (изготовитель) СИИ |
| 2 | Регистрация/сертификация | Специалисты аккредитованных испытательных лабораторий (центров сертификации) |
| 3 | Внедрение, сопровождение,  эксплуатация | Пользователи,  специалисты обслуживающей организации |

**7 Общие принципы и порядок оценки качества СИИ**

**7.1 Разработка программы тестирования (испытаний)**

Программа тестирования (испытаний) определяет цели, алгоритм тестирования (испытаний) СИИ и методологию оценки полученных результатов, порядок и содержание мониторинга, ведения записей о ходе тестирования.

**7.1.1 Общие требования к программе тестирования (испытаний)**

7.1.1.1 Программа тестирования (испытаний) разрабатывается в соответствии с описанием назначения СИИ, результатами оценки рисков и анализом данных, касающихся безопасности, существенных характеристик и субхарактеристик СИИ в соответствии с ее назначением и предполагаемым методом применения.

7.1.1.2 При разработке программ и методик тестирования (испытаний) СИИ и определении номенклатуры показателей качества и критериев оценки должны использоваться действующие отраслевые, национальные и международные стандарты (например, ГОСТ Р 52633.1-2010 и ГОСТ Р 52633.2-2010).

7.1.1.3 Программа тестирования должна быть разработана таким образом, чтобы полученные результаты тестирования (испытаний) позволили оценить, подходит ли исследуемая СИИ для цели(ей) предназначения. Программа тестирования должна быть разработана таким образом, чтобы обеспечить надежность, валидность и репрезентативность результатов тестирования.

7.1.1.4 В программе тестирования (испытаний) должны быть четко определены гипотеза и цели (первичные и вторичные), а также требования к тестовому набору данных, заявленные показатели качества и критерии оценки.

7.1.1.5 При подготовке программы тестирования (испытаний) должен быть проведен объективный анализ доступных научных данных, позволяющий обосновать полноту и достаточность набора показателей качества, правильность критериев оценки, достаточность и репрезентативность тестовых наборов данных.

**7.1.2 Создание экспертной группы**

7.1.2.1 При создании экспертной группы по оценке качества СИИ требуется:

- при выборе состава экспертной группы обеспечить ее представительность;

- определить способ организации работы с экспертами и метод формирования экспертных оценок;

- выбрать методику обработки оценок группы экспертов и критерий (-ии) согласованности экспертных мнений.

7.1.2.2 Эксперты, принимающие участие в тестировании (испытании) и последующей оценке полученных результатов на этапах разработки и оценки соответствия СИИ, должны обладать профессиональными знаниями и компетенциями, подтвержденными соответствующими документами.

7.1.2.3 Полученные результаты работы экспертной группы должны обеспечивать приемлемую сходимость полученных оценок, в противном случае требуется пересмотреть методику проведения экспертных оценок.

**7.1.3 Определение требований к качеству**

7.1.3.1. Для количественного измерения показателя качества (характеристики или субхарактеристики) используются метрики.

7.1.3.2. Метрика качества mi позволяет определить меру близости значения некоторой существенной характеристики оцениваемой системы искусственного интеллекта значению соответствующей характеристики некоторой эталонной (референтной) системы

7.1.3.3. Необходимым условием соответствия характеристики заданным критериям качества является выполнение требования (рис. 2):

где – измеренное значение i-ой характеристики;

– установленное значение i-ой характеристики;

– допустимые отклонения i-ой характеристики оцениваемой СИИ от установленного значения в направлении убывания и возрастания соответственно.

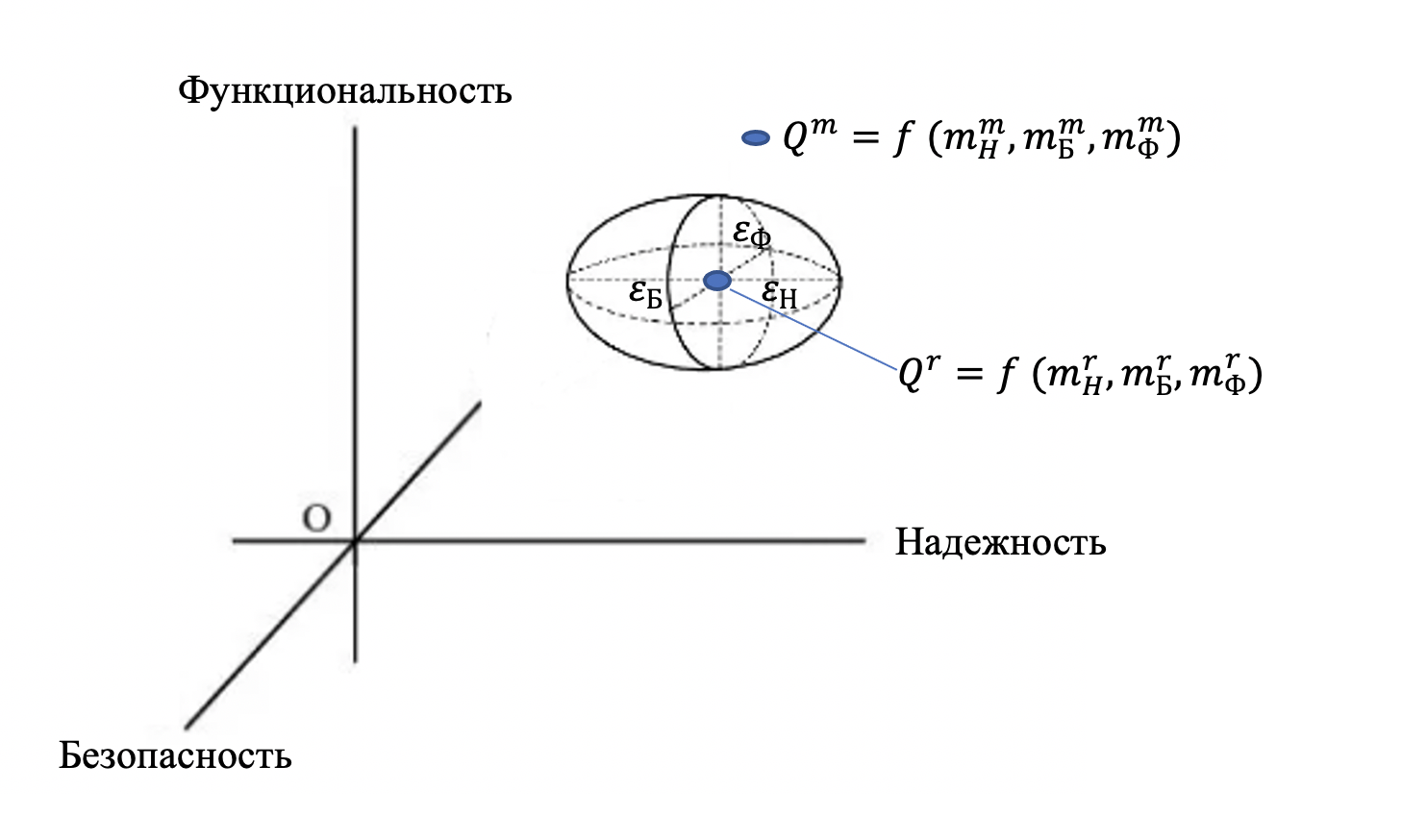


Рисунок 2 – Пример несоответствия качества СИИ по критерию функциональности (надежность и безопасность соответствуют установленным критериям качества и соответственно).

7.1.3.4. Критерии качества отражают потребности конечного пользователя СИИ и поэтому должны быть определены на этапе подготовки технического задания и задокументированы производителем в соответствующей технической и эксплуатационной документации.

**7.1.4 Выбор и обоснованность существенных характеристик.**

7.1.4.1. Выбор номенклатуры существенных характеристик (субхарактеристик) для конкретной СИИ осуществляется с учетом ее назначения и требований областей применения (см. таблицу 2).

7.1.4.2. Выбранная номенклатура существенных характеристик (субхарактеристик) и соответствующих им метрик, а также критериев оценки фиксируется в техническом задании на разработку СИИ.

Таблица 2 – Существенные характеристики и субхарактеристики СИИ для модели качества продукта

|  |  |  |
| --- | --- | --- |
| Характеристика | Субхарактеристика  1-ого рода | Субхарактеристика  2-ого рода |
| Функциональность | Функциональная пригодность (functionality) | - функциональная полнота (functional completeness)  - функциональная корректность (точность) (functional correctness);  - функциональная целесообразность (functional appropriateness);  - способность к самообучению (ability to learn) |
| Функциональная производительность (performance efficiency) | - временная эффективность (time behaviour);  - ресурсоемкость (resource utilization);  - производительные возможности (capacity) |
| Совместимость (compatibility) | **-** соответствие (co-existence);  - функциональная совместимость (interoperability);  - контролируемость (controllability) |
| Мобильность (portability) | - адаптируемость (adaptability);  - простота внедрения (installability) |
| Практичность (usability) | - определимость пригодности (appropriateness recognizability);  - изучаемость (learnability);  - управляемость (operability);  - защищенность от ошибки пользователя (user error protection);  - эстетика пользовательского интерфейса (user interface aesthetics);  - доступность (accessibility);  - взаимодействие (collaborability);  - объяснимость (explainability) |
| Сопровождаемость (maintainability) | - модульность (modularity);  - анализируемость (analysability);  - изменяемость (modifiability);  - тестируемость (testability);  - настриваемость (evolution). |
| Надежность | Надежность (reliability) | - завершенность (maturity);  - отказоустойчивость (fault tolerance);  - восстанавливаемость (recoverability);  - устойчивость (робастность) (robustness); |
| Безопасность | Защищенность (security) | - конфиденциальность (confidentiality);  - целостность (integrity);  - неотказуемость (non-repudiation);  - подотчетность (accountability);  - подлинность (authenticity);  - приватность (неприкосновенность частной жизни) (privacy) |

7.1.4.3. При выборе представительного набора существенных характеристик и субхарактеристик СИИ целесообразно руководствоваться следующими принципами (см. ГОСТ Р ИСО/МЭК 9126-93):

- полнота (достаточность) набора характеристик для принятия решения о возможности использования СИИ при решении конкретной прикладной задачи;

- простота и возможность оценки характеристик путем установления соответствующей (-их) метрик(и), ее (их) уровней ранжирования и оценки;

- простота и возможность измерения значений характеристик;

- отсутствие дублирования (перекрытия диапазонов) между используемыми характеристиками;

- соответствие установившимся понятиям и терминологии;

- возможность последующего уточнения и детализации характеристик.

**7.2 Алгоритм оценки качества СИИ**

Процедура оценки качества СИИ должна включать три основных этапа: подготовительные работы, тестирование и оценка показателей качества, анализ и интерпретация результатов тестирования.

**7.2.1 Подготовительные работы**

7.2.1.1 Подготовительные работы включают:

- установление целей и задач тестирования (испытаний);

- определение испытательного стенда и оценка условий проведения тестирования;

- выбор и обоснование набора существенных характеристик и метрик их оценки;

- выявление и определение диапазона изменений значимых, наиболее существенных факторов (внешних воздействий), оказывающих влияние на работу СИИ;

- составление методики проведения испытаний и подготовка программы тестирования (испытаний) с указанием состава экспертной группы;

7.2.1.2 Перед проведением оценки качества СИИ необходимо удостовериться в отсутствии существенных различий между средой проведения тестирования и средой эксплуатации, т.е. убедиться, что потенциальные различия не влияют на надежность, валидность и репрезентативность результатов тестирования. Примеры представлены в таблице 3.

7.2.1.3 При создании требуемых условий тестирования необходимо выделить значимые, наиболее существенные факторы (внешние воздействия), оказывающие влияние на работу СИИ. Для каждого существенного фактора требуется установить диапазон возможных изменений (закон распределения) с целью воспроизведения во время тестирования СИИ.

7.2.1.4 Задача выявления значимых, наиболее существенных факторов является одной из важных при классификации условий эксплуатации и решается путем:

- сбора априорной информации и проведения анализа имеющихся теоретических и экспериментальных данных о внешних воздействиях и режимах функционирования СИИ в реальных условиях применения;

- определения функциональной связи между воздействующими факторами (в т.ч. параметрами нагруженности) и характеристиками (субхарактеристиками).

Таблица 3 – Примеры различий между условиями тестирования и эксплуатации

|  |  |  |
| --- | --- | --- |
| № п/п | Потенциальные различия | Примеры |
| 1 | Различия между средой проведения тестирования и средой эксплуатации | - характеристики работы вычислительной в тестовой среде выше/ сопоставимы/ниже соответствующих характеристик в среде эксплуатации;  - производительность интерфейса пользователя тестовой среды выше/ сопоставима/ниже производительности интерфейса среды эксплуатации;  - производительность ЦПУ тестового компьютера выше/сопоставима/ниже производительности ЦПУ эксплуатационного компьютера;  - климатические условия, условия освещенности, различные помеховые факторы, которые оказывают влияние на эффективность работы систем, использующих датчики, основанные на различных физических принципах;  - и т.п. |
| 2 | Различия между тестированием выполнения и фактическим эксплуатационным выполнением | - охват функциональных возможностей;  - количество и репрезентативность наборов данных;  - автоматизированное испытание в режиме реального времени;  - стрессовые загрузки;  - 24 час 7 дней в неделю (без перерыва) работа;  - уместность данных для испытания исключений и ошибок;  - использование ресурса;  - уровни прерывания  - и т.п. |
| 3 | Различия между профилями пользователя при тестировании и эксплуатационными профилями пользователя | - уровни навыков пользователей;  - пользователи-специалисты или средние пользователи;  - ограниченная группа пользователей или общественные пользователи. |

**7.2.2 Тестирование и оценка показателей качества**

7.2.2.1 Для каждой характеристики из набора существенных характеристик СИИ определяются метрики, позволяющие количественно оценить качество СИИ на определенной стадии жизненного цикла. Интегральный (-ые) показатель (-и) качества, используемые в процессе разработки, должны быть соотнесены с соответствующими показателями качества пользователя.

7.2.2.2 Для каждого метрики и показателя качества должна быть установлена шкала и критерии оценки (например, базовое значение и предельно допустимое отклонение).

7.2.2.3 После проведения тестирования полученное значение метрики (показателя качества) сравнивается с базовым значением. Заключение о соответствии принимается при выполнении следующего условия:

где – полученное значение i-ого показателя качества;

– базовое значение i-ого показателя качества;

– допустимое отклонение i-характеристик оцениваемой системы искусственного интеллекта

7.2.2.4 Вычисление интегрального показателя качества выполняется в следующем порядке:

Для каждой j-й метрики mj определяется весовой коэффициент . Сумма

весовых коэффициентов всех метрик, относящихся к одной и той же подхарактеристике должна быть равна 1:

где L – количество метрик для конкретной субхарактеристики.

Далее проводится оценка каждой i-oй субхарактеристики сi k-ой характеристики. Для этого используется следующая формула:

Для каждой i-й субхарактеристики сi определяется весовой коэффициент . Сумма весовых коэффициентов всех подхарактеристик, относящихся к одной и той же характеристике постоянна и равна 1.

где N – количество субхарактеристик конкретной k-ой характеристики.

Далее проводится оценка каждой k-ой характеристики:

Для каждой k-ой характеристики (функциональность, безопасность, надежность) определяется соответствующий коэффициент . Сумма весовых характеристик поятоянна и равна 1.

где O – количество характеристик, используемых для оценки качества СИИ.

Интегральная оценка качества Q СИИ рассчитывается по формуле

Интегральная оценка качества СИИ Q принимает значения на интервале от 0 до 1, причем, чем ближе к 1, тем выше качество СИИ.

**8. Представительный набор существенных характеристик и показатели качества СИИ.**

Приведенные в данном разделе метрики оценки качества СИИ носят рекомендательный характер.

**8.1 Общие положения**

8.1.1. В стандарте рассмотрены базовые метрики для оценки соответствующих характеристик (подхарактеристик) СИИ, а также представлены метрики точности, предлагаемые к применению в зависимости от типа решаемых задач. Однако данный набор метрик не является исчерпывающим в связи с многообразием областей применения СИИ и широким спектром решаемых с их помощью задач.

8.1.2. Набор метрик формируется на этапе подготовки технического задания, а также может быть скорректирован и дополнен на последующих стадиях разработки программного продукта. Для данных целей требуется использовать отраслевые стандарты ИИ, регламентирующие наборы существенных характеристик прикладных СИИ, предназначенных для решения различных типовых задач в соответствующих отраслях экономики и социальной сферы. Итоговый набор существенных (значимых) характеристик системы и соответствующих им показателей качества СИИ должен быть достаточен для оценки степени соответствия СИИ для решения поставленной задачи в целях обеспечения доверия со стороны пользователей.

8.1.3. Для получения достоверных результатов оценки качества СИИ целесообразно, чтобы установленные метрики обладали следующими характерными свойствами, определяющими точность проводимых измерений согласно ГОСТ 5725-1-2002:

- достоверность (влияние случайных ошибок на результат измерения не значительно или может быть учтено в процессе анализа, то есть не может повлиять на результат измерения);

- воспроизводимость (повторное измерение метрики для той же систем, используя ту же самую шкалу ранжирования и оценки, входные данные и условия проведения тестирования, различными специалистами по оценке должно привести к тем же самым результатам в пределах соответствующей устойчивости);

- повторяемость (также сходимость результатов измерений) — близость друг к другу результатов измерений одной и той же величины, выполненных повторно одними и теми же средствами, одним и тем же методом в одинаковых условиях и с одинаковой тщательностью;

- показательность (метрики): Способность метрики идентифицировать части или продукцию программного обеспечения, которые должны быть улучшены, учитывая взвешенные результаты по сравнению с ожидаемыми.

**8.2 Функциональные возможности (functionality) СИИ. Метрики**

Для оценки функциональных возможностей СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: функциональная полнота (functional completeness), функциональная корректность (functional correctness), функциональная целесообразность (functional appropriateness), способность к самообучению (ability to learn).

8.2.1 Метрики функциональной полноты (functional completeness) используются для оценки степени покрытия совокупностью функций СИИ всех определенных задач и целей пользователя в условиях отсутствия предвзятости (необъективности) СИИ.

Таблица 4 – Примеры метрик оценки функциональной полноты

|  |  |  |
| --- | --- | --- |
| Наименование метрики | Формула | Измеряемый диапазон |
| Полнота реализации функций | где А – количество недостающих или неправильно реализованных функций, обнаруженных при оценивании;  B – количество функций, описанных в технической и эксплуатационной документации. |  |

8.2.2 Метрики функциональной корректности (functional correctness) используются для оценки обеспечения CИИ степени точности результатов, а также частоты встречаемости ошибок и недопустимых отклонений.

Таблица 5 – Примеры метрик оценки функциональной корректности

|  |  |  |
| --- | --- | --- |
| Наименование метрики | Формула | Измеряемый диапазон |
| Результативность | Где А – количество результатов с отличным от требуемого уровня точности;  В – общее количество результатов |  |
| Точность вычислений | Примеры вычисления метрик для типовых задач ИИ представлены в пп. 8.2.2.1 – 8.2.2.3 | - |

8.2.3 Метрики функциональной целесообразности (functional appropriateness) используются для оценки степени функционального упрощения выполнения определенных задач и достижения целей. Например, для решения задачи пользователю предоставляется возможность выполнять только необходимые шаги, исключая любые ненужные.

Таблица 7 – Примеры метрик оценки функциональной целесообразности

|  |  |  |
| --- | --- | --- |
| Наименование метрики | Формула | Измеряемый диапазон |
| Степень автоматизации | где А – количество шагов, выполняемых СИИ без привлечения пользователя, при реализации конкретной процедуры;  В – общее количество шагов при выполнении заданной процедуры |  |

8.2.4 Метрики способности к самообучению (ability to learn) используются для оценки уровня владения СИИ умением автоматичес­ки извлекать знания из накопленного опыта и применять их для решения поставленных задач.

**8.3 Уровень производительности (performance efficiency). Метрики**

При тестировании рекомендуется проводить оценки более длительных по времени операций или рассматривать распределение времени для нескольких случаев, так как значения метрик производительности подвержены сильному влиянию условий применения, как, например, загрузка обрабатываемых данных.

В протокол тестирования следует включить параметры, влияющие на показатели качества: параметры ЦПУ, объем памяти, сетевой трафик и прочее.

Для оценки уровня производительности СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: временная эффективность (time behaviour), ресурсоемкость (resource utilization), производительные возможности (capacity).

8.3.1 Метрики временной эффективности (time behaviour) используют для оценки степени соответствия требованиям временных ресурсов, затрачиваемых пользователем в целях обеспечения точной и полной реализации конкретных задач.

Таблица 8 – Примеры метрик оценки временной эффективности

|  |  |  |
| --- | --- | --- |
| Наименование метрики | Формула | Измеряемый диапазон |
| Отклонение времени отклика | где Ti– время отклика i-ого измерения;  N – количество измерений;  Tн– допустимое время отклика, представленного в технической документации |  |
| Производительность | где А – количество однотипных задач, выполненных СИИ за время T. |  |

8.3.2 Метрики ресурсоемкости (resource utilization) используют для оценки степени удовлетворения требований по потреблению объемов ресурсов СИИ при выполнении ее функций.

8.3.3. Метрики производительных возможностей (capacity) используются для оценки степени соответствия требованиям предельных значений параметров СИИ, как, например: количество параллельно обрабатываемых наборов данных, количество параллельно работающих пользователей, емкость канала, пропускная способность по транзакциям и прочее.

**8.4. Совместимость (Compatibility). Метрики**

Способность СИИ обмениваться информацией с другими продуктами, системами или компонентами, и/или выполнять требуемые функции при совместном использовании одних и тех же аппаратных средств или программной среды.

Для оценки совместимости СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: соответствие (co-existence), функциональная совместимость (interoperability), контролируемость (controllability).

8.4.1. Метрики соответствия (co-existence) используются для оценки способности СИИ совместно функционировать с другими независимыми системами в общей среде с разделением общих ресурсов и без отрицательного влияния на любой другой продукт.

8.4.2. Метрики функциональной совместимости (interoperability) используются для оценки способности СИИ обмениваться информацией с другими системами, продуктами или компонентами и использовать такую информацию.

8.4.3. Метрики контролируемости (controllability) используются для оценки степени управляемости СИИ.

**8.5 Мобильность (portability)**

Для оценки мобильности СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: адаптируемость (adaptability), простота внедрения (installability).

8.5.1 Метрики адаптируемости (adaptability) используются для оценки усилий пользователя СИИ, направленных на адаптацию для потребностей пользователя предварительно представленного ПО.

8.5.2 Метрики простоты внедрения (installability) используются для оценки простоты эффективной и рациональной, успешной установки и/или удаления СИИ в заданной среде.

**8.6. Практичность (usability) СИИ**

Для оценки практичности СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: - определимость пригодности (appropriateness recognizability), изучаемость (learnability), управляемость (operability), защищенность от ошибки пользователя (user error protection), эстетика пользовательского интерфейса (user interface aesthetics), доступность (accessibility), взаимодействие (collaborability), объяснимость (explainability).

8.6.1 Метрики определимости пригодности (appropriateness recognizability) используются для оценки соответствия СИИ функциональным потребностям пользователя, включая демонстрационные материалы, обучающие программы, документацию и прочее.

8.6.2 Метрики изучаемости (learnability) используются для оценки достижения СИИ конкретных целей обучения для эксплуатации СИИ с требуемой эффективностью, результативностью, свободой от риска и в соответствии с требованиями в указанном контексте использования.

8.6.3 Метрики управляемости (operability) используются для оценки наличия в СИИ атрибутов, обеспечивающих простое управление и контроль за ее функционированием.

8.6.4 Метрики защищенности от ошибки пользователя (user error protection) используются для оценки степени готовности СИИ предотвратить ошибки оператора, которые могут привести к сбою технических и программных средств СИИ, а также искажению, уничтожению, копированию, блокированию доступа к информации и прочее.

8.6.5 Метрики эстетики пользовательского интерфейса (user interface aesthetics) применяются для оценки степени удовлетворенности пользователя интерфейсом СИИ взаимодействия с пользователем.

8.6.6 Метрики доступности (accessibility) используются для оценки возможности использования СИИ широким кругом людей с самыми разными возможностями.

Доступность для людей с ограниченными возможностями может быть задана или измерена либо как степень, в которой СИИ может быть применена пользователями с указанными ограниченными возможностями для достижения определенных целей с эффективностью, результативностью, свободой от риска и в соответствии с требованиями в указанном контексте использования, либо как наличие свойств продукта для поддержки доступности.

8.6.7 Метрики взаимодействия (collaborability) используются для оценки надежности контроля и управления потоками данных («оркестрирование», совместное управление) между пользователями различных групп.

8.6.8 Метрики объяснимости (explainability) используются для оценки понятности и прозрачности процесса принятия решений ИИ для пользователя .

**8.7 Сопровождаемость (maintainability) СИИ. Метрики**

Для оценки сопровождаемости СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: модульность (modularity), анализируемость (analysability), изменяемость (modifiability), тестируемость (testability), настриваемость (evolution).

8.7.1 Метрики анализируемости (analysability) используются для оценки усилий пользователя или специалиста по сопровождению СИИ или затрат ресурсов при попытке обнаружить ошибки или причины отказов, или для определения блоков, которые необходимо изменить.

8.7.2 Метрики изменяемости (modifiability) используются для оценки степени простоты эффективной и рациональной модификации СИИ без снижения качества работы СИИ.

8.7.3 Метрики тестируемости (testability) используются для оценки простоты выполнения тестирования СИИ с целью определения соответствия заданным критериям.

8.7.4 Метрики настраиваемости (evolution) используется для определения степени надежности мониторинга СИИ дрейфа данных, вычисления веса важности и правильности принятия меры для переобучения или перестройки модели.

**8.8 Надежность (reliability) СИИ. Метрики**

Для оценки надежности СИИ рекомендуется применять следующий набор субхарактеристик 2-ого рода: завершенность (maturity), отказоустойчивость (fault tolerance), восстанавливаемость (recoverability), устойчивость (робастность) (robustness).

8.8.1 Метрики завершенности (maturity) используются для оценки независимости СИИ от отказов, существующих в самой СИИ.

Таблица 9 – Примеры метрик завершенности

|  |  |  |
| --- | --- | --- |
| Наименование метрики | Формула | Измеряемый диапазон |
| Плотность отказов по отношению к тестовым наборам | где O – количество обнаруженных отказов СИИ;  T – количество выполненных тестовых наборов |  |
| Устранение ошибок | где А – количество исправленных ошибок  B – общее количество реально обнаруженных (или спрогнозированных) ошибок |  |
| Тестовое покрытие | где А – количество фактически выполненных тестовых примеров, которые отображают сценарий эксплуатации СИИ,  B – количество тестовых примеров, которые необходимо выполнить, чтобы удовлетворить требованиям, определяющим цели и условия эксплуатации СИИ |  |
| Коэффициент аварийных отказов | где А – количество аварийных отказов СИИ,  B – количество отказов СИИ |  |

8.8.2 Метрики отказоустойчивости (fault tolerance) используются для оценки способности СИИ поддерживать определенный уровень производительности в случаях возникновения ошибок эксплуатации.

8.8.3 Метрики восстанавливаемости (recoverability) используются для оценки способности СИИ восстанавливать отвечающий требованиям уровень производительности, а также данные, подвергаемые прямому действию в случае отказа.

8.8.4 Метрики устойчивости (робастности) (robustness) используются для оценки способности СИИ обеспечивать требуемую точность (прецизионность и правильность) выходных данных при наличии различного рода выбросов, помех.

**9 Набор входных и выходных данных**

**9.1 Требования к набору входных и выходных данных для тестирования СИИ**

9.1.1 Для целей тестирования в процессе жизненного цикла СИИ применяются следующие наборы данных (см. таблицу 10):

- базовый демонстрационный набор данных – образцовый аннотированный набор данных, сформированный в соответствии с нормативно-техническим документом, устанавливающим унифицированные требования к проведению испытаний СИИ определённого типа;

- дополнительный демонстрационный набор данных – дополнительный аннотированный набор данных, предоставляемый заказчиком СИИ при формировании уточнённых требований к системе с учётом конкретного функционала и условий ее эксплуатации;

- полный демонстрационный набор данных – совокупность базового и дополнительного контрольных наборов данных;

- обучающий набор данных – набор данных, формируемый на основе демонстрационного набора данных и необходимый для создания СИИ. При формировании обучающего набора данных широко применяются технологии аугментации данных, позволяющие повысить качество создаваемой СИИ;

- тестовый набор данных – набор данных, формируемый в органах по оценке соответствия на основе демонстрационного набора данных и необходимый для проведения сертификации (регистрации), испытаний или аттестации СИИ.

Таблица 10 – Типы наборов данных

|  |  |  |
| --- | --- | --- |
| Тип набора данных | Назначение | Способ формирования |
| Базовый  демонстрационный | Определение минимальных требований к  функциональным  характеристикам СИИ | В соответствии с нормативно-техническим документом (частным стандартом), определяющим порядок оценки функциональных характеристик СИИ и иллюстрирующим требования к демонстрационному набору данных |
| Дополнительный  демонстрационный | Уточнение требований к функциональным характеристикам СИИ | Предоставляется заказчиком при формировании уточнённых требований к СИИ с учётом конкретного функционала и условий ее эксплуатации |
| Полный  демонстрационный | Формирование исходного образца для подготовки обучающего набора данных | В результате объединения базового и дополнительного демонстрационных наборов данных |
| Обучающий | Обучение СИИ | На основе демонстрационного набора данных с учетом выполнения требования статистической эквивалентности демонстрационному набору данных |
| Тестовый | Оценка соответствия СИИ | На основе демонстрационного набора данных в соответствии с требованиями сертификации (регистрации), испытаний или аттестации СИИ |

9.1.2. При формировании тестового набора данных в целях обеспечения качества данных по ИСО/МЭК 25012 необходимо учитывать следующие требования к тестовым наборам данных:

- представительность: отклонение оценок функциональных характеристик СИИ, полученных при проведении испытаний на тестовом наборе данных, от апостериорных значений этих характеристик, полученных в результате эксплуатации СИИ в предусмотренных условиях эксплуатации, с заданной вероятностью не должно превышать некоторую определённую, допустимо малую величину. (К критериями представительности тестового набора относят соответствие статистических характеристик и сложности решения задачи на тестовом наборе данных в предусмотренных условиях эксплуатации);

- безызбыточность: при формировании тестового набора данных необходимо стремиться к тому, чтобы его объём был минимально достаточным для выполнения требования представительности;

- объективность: тестовый набор данных должен доказуемо принадлежать к той же генеральной совокупности, что и демонстрационный набор данных. При необходимости доказательства объективности тестового НД должны быть предоставлены разработчику и другим заинтересованным сторонам без нарушения конфиденциальности тестового НД;

- конфиденциальность: в отношении тестового набора данных органами по оценке соответствия должен быть обеспечен режим конфиденциальности, исключающий использование тестового набора данных разработчиком на стадии создания СИИ. При этом должна быть обеспечена объективность тестового набора данных.

|  |  |  |
| --- | --- | --- |
|  | | |
| УДК 615.841:006.354 |  | ОКС 11.040.01 | |
| Ключевые слова: системы искусственного интеллекта, контроль качества, метрики | | |
|  | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Руководитель  разработки |  |  |  |  | К.А. Сергунова |
|  | должность |  | личная подпись |  | инициалы, фамилия |