ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 59898— 2021

ОЦЕНКА КАЧЕСТВА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Общие положения

Издание официальное

Москва
Российский институт стандартизации
2021

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Федеральным государственным автономным образовательным учреждением высшего образования «Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ)

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 26 ноября 2021 г. 1620-ст

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 (*gost.e-ecolog.ru

© Оформление. ФГБУ «РСТ», 2021

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

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Термины и определения 2
4 Сокращения 4
5 Модель качества 4
6 Методология и критерии оценки качества на стадиях жизненного цикла СИИ 5
7 Общие принципы и порядок оценки качества СИИ 6
7.1 Разработка программы тестирования (испытаний) 6
7.2 Алгоритм оценки качества СИИ 8
8 Представительный набор существенных характеристик и показатели качества СИИ 11
8.1 Общие положения 11
8.2 Функциональные возможности (functionality) СИИ. Метрики 11
8.3 Уровень производительности (performance efficiency) СИИ. Метрики 15
8.4 Способность СИИ к взаимодействию (compatibility). Метрики 15
8.5 Мобильность (portability) СИИ. Метрики 16
8.6 Практичность (usability) СИИ. Метрики 16
8.7 Сопровождаемость (maintainability) СИИ. Метрики 16
8.8 Надежность (reliability) СИИ. Метрики 17
9 Требования к набору входных и выходных данных для тестирования СИИ 18
Библиография 20

Введение

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

- определить текущие параметры СИИ и выполнить действия, направленные на повышение ее надежности, производительности, востребованности и целенаправленности, 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 снижению уровня 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 ГОСТ Р ИСО/МЭК 25010—2015, 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 ГОСТ Р 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 качества СИИ; приведена классификация существенных характеристик и показателей качества СИИ.

ОКС 11.040.01

Поправка к ГОСТ Р 59898—2021 Оценка качества систем искусственного интеллекта. Общие положения

В каком месте	Напечатано	Должно быть
Библиографические данные	ОКС 11.040.01	ОКС 35.020

(ИУС № 9 2025 г.)

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ОЦЕНКА КАЧЕСТВА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Общие положения

Quality assurance of artificial intelligence systems. General

Дата введения — 2022—03—01

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 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 СИИ, приведенные в настоящем стандарте, могут быть дополнены требованиями конкретных стандартов в соответствующих областях применения СИИ (например, для оценки биометрических технологий и систем необходимо использовать ГОСТ Р ИСО/МЭК 19795-1, ГОСТ Р 58292, ГОСТ Р ИСО/МЭК 19795-6, ГОСТ Р 58624.1, ГОСТ Р 58667.3).

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), для решения конкретных практически значимых задач. Настоящий стандарт не может быть использован для систем «сильного» или «общего» искусственного интеллекта.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ 34.601 Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Стадии создания

ГОСТ Р 8.673 Государственная система обеспечения единства измерений. Датчики интеллектуальные и системы измерительные интеллектуальные. Основные термины и определения

ГОСТ Р 8.734 Государственная система обеспечения единства измерений. Датчики интеллектуальные и системы измерительные интеллектуальные. Методы метрологического самоконтроля

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

ГОСТ Р ИСО/МЭК 19795-1 Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 1. Принципы и структура

ГОСТ Р ИСО/МЭК 19795-6 Информационные технологии. Биометрия. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 6. Методология проведения оперативных испытаний

Издание официальное

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

ГОСТ Р 58292 (ИСО/МЭК 19795-2:2007) Информационные технологии. Биометрия. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 2. Методы проведения технологического и сценарного испытаний

ГОСТ Р 58624.1 Информационные технологии. Биометрия. Обнаружение атаки на биометрическое предъявление. Часть 1. Структура

ГОСТ Р 58667.3 Информационные технологии. Биометрия. Пиктограммы, значки и символы для использования в биометрических системах. Часть 3. Приложения, осуществляющие работу с изображениями лиц

ГОСТ Р 58450 Изделия медицинские с измерительными функциями. Контроль состояния

ГОСТ Р 59276—2020 Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения

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

3.2 **аугментация данных** (data augmentation): Процесс создания дополнительного набора данных из имеющегося набора данных.

Примечание — Применяется для увеличения обучающего набора данных путем модификации существующего набора данных.

3.3 **базовое значение показателя качества** (baseline quality score): Значение показателя качества системы искусственного интеллекта, принятое за основу при сравнительной оценке ее качества.

3.4 **безопасность** (safety): Свойство системы искусственного интеллекта сохранять состояние, характеризующееся отсутствием недопустимого риска, при использовании ее по назначению в условиях, предусмотренных изготовителем.

3.5 **выборка** (sample): Набор данных, представляющий собой подмножество генеральной совокупности.

3.6

выброс (outlier): Элемент маломощного подмножества выборки, существенно отличающийся от остальных элементов выборки.

Примечания

1 Классификация наблюдения или подмножества выборки как выброс (или выбросы) зависит от выбранной модели генеральной совокупности, из которой отобрана выборка. Выброс не рассматривают как истинный элемент генеральной совокупности.

2 Выброс может появиться из другой генеральной совокупности, быть результатом некорректной регистрации данных или общей ошибкой измерений.

3 Подмножество может содержать одно или несколько наблюдений.

[ГОСТ Р ИСО 16269-4—2017, статья 2.2]

3.7 **генеральная совокупность** (general sample): Репрезентативное множество всех возможных прецедентов.

3.8

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, статья 3.6]

3.9 **качество** (quality): Совокупность характеристик и свойств СИИ, обусловливающих ее способность удовлетворять установленным или предполагаемым требованиям в соответствии с ее назначением.

3.10 **критерий оценки качества** (quality assessment criterion): Набор определенных и задокументированных правил и условий, которые используются для решения о приемлемости общего качества конкретной СИИ.

3.11

метаданные (metadata): Данные о данных или элементах данных, которые могут включать описание, а также сведения о владельце данных, путях доступа к ним, правах доступа и изменчивости данных.

[ГОСТ Р ИСО/МЭК 20546—2021, статья 3.1.24]

3.12 **метрика** (metric): Материальная мера некоторых аспектов характеристик качества.

3.13 **набор данных** (dataset): 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.

3.14

обогащенные данные: Данные, объединенные или дополненные логически связанными данными, полученными от других поставщиков данных, а также данные, полученные в результате объединения по различным признакам и категориям.

[ГОСТ 59237—2020, статья 18]

3.15

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.

[ГОСТ Р 59276—2020, статья 3.9]

3.16 **оценка качества** (quality assurance): Совокупность операций, включающих выбор номенклатуры показателей качества, определение значений этих показателей и сравнение их с базовыми значениями.

3.17

предвзятость, необъективность (bias): Свойство системы искусственного интеллекта, заключающееся в принятии ошибочных решений, связанных со статистической смещенностью обучающей выборки исходных данных или необъективностью применяемых при ее создании результатов экспертной работы.

[ГОСТ Р 59276—2020, статья 3.11]

3.18

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.

[ГОСТ Р 59276—2020, статья 3.13]

3.19 **прецедент** (use case): Спецификация последовательности действий, включая опции, которые СИИ может выполнить при взаимодействии с пользователями.

3.20

система искусственного интеллекта: Техническая система, в которой используются технологии искусственного интеллекта, и обладающая искусственным интеллектом.

[ГОСТ Р 59276—2020, статья 3.16]

3.21

существенные (значимые) характеристики системы искусственного интеллекта: Характеристики системы искусственного интеллекта, определяющие ее 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.

Примечание — Характеристики систем искусственного интеллекта, подтверждение соответствия которых установленным требованиям может быть выполнено исключительно разработчиком системы, не относятся к существенным.

[ГОСТ Р 59276—2020, статья 3.17]

4 Сокращения

В настоящем стандарте применены следующие сокращения:

ЖЦ — жизненный цикл;

ИИ — искусственный интеллект;

НД — набор данных;

ПО — программное обеспечение;

СИИ — система искусственного интеллекта;

ЦПУ — центральное процессорное устройство.

5 Модель качества

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

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

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

5.4 Для оценки качества СИИ могут быть использованы 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 как продукта.

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

$$F(S) = \{f_1, f_2, \ldots f_N\}, \quad (1)$$

где S — оцениваемая СИИ;

F — представительный набор существенных характеристик;

f_i — i-я существенная характеристика.

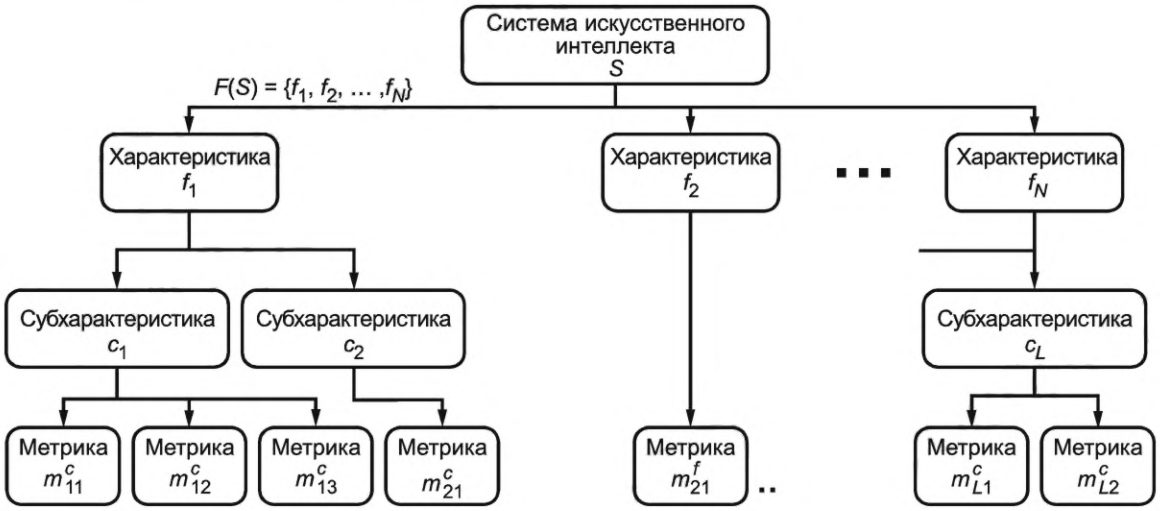


Рисунок 1 — Дерево существенных характеристик и метрик СИИ

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

6 Методология и критерии оценки качества на стадиях жизненного цикла СИИ

6.1 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 (техническое задание, технический проект, рабочая документация);
- контроле качества в процессе изготовления СИИ;
- проверке эффективности модификации СИИ на стадии сопровождения;
- контроле состояния СИИ, включая поверку и калибровку в случае отнесения ее к СИИ с измерительными функциями (например, по ГОСТ Р 8.734 и ГОСТ 58450).

6.2 В зависимости от стадии жизненного цикла тестирование СИИ могут выполнять различные группы лиц (см. таблицу 1).

Таблица 1 — Специалисты, осуществляющие тестирование СИИ

№ п/п	Стадия жизненного цикла	Специалисты, осуществляющие тестирование СИИ
1	Разработка (изготовление)	Разработчик (изготовитель) СИИ
2	Оценка соответствия (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 Программу тестирования (испытаний) разрабатывают в соответствии с описанием назначения СИИ, установленными и предполагаемыми требованиями к 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 следует использовать действующие отраслевые, национальные и международные стандарты (например, для оценки биометрических технологий и систем — ГОСТ Р ИСО/МЭК 19795-1, ГОСТ Р 58292, ГОСТ Р ИСО/МЭК 19795-6, ГОСТ Р 58624.1, ГОСТ Р 58667.3).

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 Экспертную группу по оценке качества СИИ создают для решения одной или нескольких задач:

- аннотирование (разметка) данных;
- выбор номенклатуры характеристик (субхарактеристик) и соответствующих им метрик, а также критериев оценки, включая требования к ним;
- участие в тестировании (испытании) СИИ и последующей оценке полученных результатов.

7.1.2.2 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.

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

7.1.2.4 Полученные результаты работы экспертной группы должны обеспечивать приемлемую сходимость 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 m_j позволяет определить меру близости значения некоторой существенной характеристики f_i^m оцениваемой системы искусственного интеллекта S^m значению соответствующей характеристики f_i^r некоторой эталонной (референтной) системы S^r.

7.1.3.3 Необходимым условием соответствия характеристики заданным критериям качества является выполнение требования (рисунок 2)

$$m_i^m \in \left[m_i^r - \varepsilon_{i-}, m_i^r + \varepsilon_{i+} \right], \quad (2)$$

где m_i^m — измеренное значение i-й характеристики;

m_i^r — установленное значение i-й характеристики;

ε_{i-}, ε_{i+} — допустимые отклонения i-й характеристики оцениваемой СИИ S^m от установленного значения в направлении убывания и возрастания, соответственно.

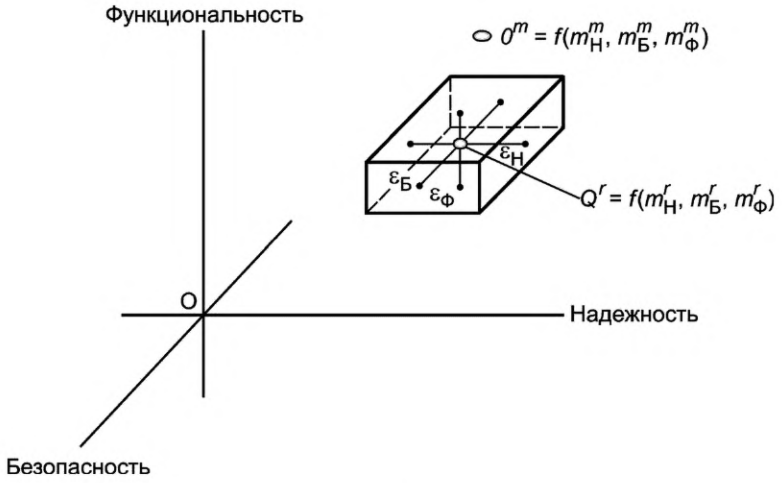


Рисунок 2 — Пример несоответствия качества СИИ по критерию функциональности $\varepsilon_Ф$ (надежность и безопасность соответствуют установленным критериям качества с допустимыми отклонениями $\varepsilon_Н$ и $\varepsilon_Б$)

7.1.3.4 Критерии качества отражают потребности конечного пользователя СИИ, и поэтому должны быть определены на этапе подготовки технического задания и задокументированы разработчиком 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, включая требования к ним, фиксируется в техническом задании на разработку СИИ.

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

Группа характеристик	Характеристика по ГОСТ Р 59276	Существенная характеристика	Субхарактеристика
Функциональность	Функциональные возможности	Функциональные возможности (functionality)	Функциональная пригодность (functional appropriateness); функциональная корректность (правильность) (functional correctness); согласованность (compliance); функциональная полнота (functional completeness); способность к самообучению (ability to learn)
		Способность к взаимодействию (compatibility)	Соответствие (co-existence); функциональная совместимость (interoperability); контролируемость (controllability)
	Эффективность	Уровень производительности (performance efficiency)	Характер изменения во времени (time behaviour); характер изменения (использования) ресурсов (resource utilization); производительные возможности (capacity)

Окончание таблицы 2

Группа характеристик	Характеристика по ГОСТ Р 59276	Существенная характеристика	Субхарактеристика
Функциональность	Мобильность	Мобильность (portability)	Адаптируемость (adaptability); простота внедрения (installability); (взаимо)заменяемость (replacebility)
	Практичность	Практичность (usability)	Понятность (explainability); изучаемость (learnability); простота использования (operability); защищенность от ошибки пользователя (user error protection); эстетика пользовательского интерфейса (user interface aesthetics); доступность (accessibility); взаимодействие (collaborability)
	Сопровождаемость	Сопровождаемость (maintainability)	Анализируемость (analysability); изменяемость (modifiability); устойчивость (tability); тестируемость (testability); модульность (modularity); настраиваемость (evolution)
Надежность	Надежность	Надежность (reliability)	Стабильность (maturity); устойчивость к ошибке (отказоустойчивость) (fault tolerance); восстанавливаемость (recoverability); 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 принципами, приведенными в разделе 5 ГОСТ Р 59276—2020:

- 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);
- выбор (создание) испытательного стенда и оценка условий проведения тестирования;
- определение набора данных для тестирования и критерия выявления выбросов в данных;
- 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 тестирования (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.

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 СИИ 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 доступным функционалом при тестировании и эксплуатации	Охват функциональных возможностей; количество и репрезентативность наборов данных; уровень стрессовых нагрузок; длительность пребывания в рабочем режиме; степень использования ресурса; и т. п.
3	Различия между профилями пользователя при тестировании и эксплуатационными профилями пользователя	Уровни навыков пользователей; уровень подготовки пользователей (пользователи-специалисты или средние пользователи); квалификация пользователей

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 том числе параметрами нагруженности) и характеристиками (субхарактеристиками).

7.2.2 Тестирование и оценка показателей качества

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

7.2.2.2 Для каждой пары «показатель качества и метрика» 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) нормируется (в случае, если используются не все метрики) таким образом, чтобы диапазон возможных значений лежал на интервале от 0 до 1, причем, чем ближе к 1, тем ближе полученное значение к базовому значению показателя качества, и равно 1 при их совпадении.

7.2.2.4 Вычисление интегрального показателя качества выполняют в следующем порядке:

Для каждой j-й метрики m_j задается весовой коэффициент v_j^m. Сумма весовых коэффициентов всех метрик, относящихся к одной и той же субхарактеристике, должна быть равна 1

$$\sum_{j=1}^{L} v_j^m = 1, \tag{3}$$

где L — количество метрик для конкретной субхарактеристики.

Далее проводят оценку каждой i-й субхарактеристики c_i k-й характеристики. Для этого используется следующая формула

$$c_i^k = \frac{\sum_{j=1}^{L1}(m_j \cdot v_j^m)}{\sum_{j=1}^{L1} v_j^m}, \tag{4}$$

где $L1$ — количество метрик, использованных при оценке конкретной субхарактеристики ($L1 \le L$).

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

$$\sum_{i=1}^{N} v_i^c = 1, \tag{5}$$

где N — количество субхарактеристик конкретной k-й характеристики.

Далее проводят оценку каждой k-й характеристики

$$f_k = \frac{\sum_{i=1}^{N1}(c_i^k \cdot v_i^c)}{\sum_{i=1}^{N1} v_i^c}, \tag{6}$$

где $N1$ — количество субхарактеристик, использованных при оценке конкретной k-й характеристики ($N1 \le N$).

Для каждой k-й характеристики (функциональность, безопасность, надежность) определяют соответствующий коэффициент v_k^Q. Сумма весовых характеристик постоянна и равна 1

$$\sum_{k=1}^{O} v_k^Q = 1, \tag{7}$$

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

Интегральную оценку качества Q СИИ рассчитывают по формуле

$$Q = \sum_{k=1}^{O}(f_k \cdot v_k^Q) Q = \sum_{k=1}^{O}(f_k \cdot v_k^Q). \tag{8}$$

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

7.2.2.5 При проведении сравнительных оценок нескольких СИИ при тестировании должны быть использованы единые перечни субхарактеристик и соответствующих им метрик, унифицированные шкалы по каждому показателю качества, а также применены единые правила нормирования по 7.2.2.3.

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. 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 формируют на этапе подготовки технического задания, он может быть скорректирован и дополнен на последующих стадиях разработки СИИ. Для данных целей требуется использовать отраслевые и иные нормативно-технические документы, 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 Р ИСО 5725-1:

- достоверность — влияние случайных ошибок на результат измерения незначительно или может быть учтено в процессе анализа, то есть не может повлиять на результат измерения;
- воспроизводимость — повторное измерение метрики для той же системы, 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 погрешности;
- 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

8.2.1 Для оценки функциональных возможностей СИИ рекомендуется применять следующий набор субхарактеристик: функциональная пригодность (functional appropriateness), функциональная корректность (правильность) (functional correctness), согласованность (compliance), функциональная полнота (functional completeness), способность к самообучению (ability to learn).

8.2.2 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 (см. таблицу 4).

Таблица 4 — Примеры метрик оценки функциональной полноты

Наименование метрики	Формула	Измеряемый диапазон
Полнота реализации функций	$M_1 = 1 - \frac{A}{B}$, (9) где A — количество недостающих или неправильно реализованных функций, обнаруженных при оценивании; B — количество функций, описанных в технической и эксплуатационной документации	$M_1 \in [0,1]$

8.2.3 Метрики функциональной корректности (правильности) (functional correctness) используются для оценки обеспечения СИИ степени точности результатов, а также частоты встречаемости ошибок и недопустимых отклонений (см. таблицу 5).

Таблица 5 — Примеры метрик оценки функциональной корректности

Наименование метрики	Формула	Измеряемый диапазон
Результативность	$$M_2 = 1 - \frac{A}{B}, \quad (10)$$ где A — количество результатов с отличным от требуемого уровнем точности; B — общее количество результатов	$M_2 \in [0,1]$
В задачах регрессии		
Средняя квадратичная ошибка	$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \quad (11)$$ где y — выходные данные СИИ; $\hat{y}$ — эталонные (референсные) выходные данные; N — количество результатов работы СИИ	$MSE \geq 0$
Средняя абсолютная ошибка	$$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert, \quad (12)$$ где y — выходные данные СИИ; $\hat{y}$ — эталонные (референсные) выходные данные; N — количество результатов работы СИИ	$MAE \geq 0$
В задачах классификации и обнаружения		
Доля правильных исходов (accuracy)	$$A = \frac{TP + TN}{TP + TN + FP + FN}, \quad (13)$$ где TP — количество истинно положительных исходов; TN — количество истинно отрицательных исходов; FP — количество ложно положительных исходов; FN — количество ложно отрицательных исходов	$A \in [0,1]$
Точность (precision, relevance)	$$Pr = \frac{TP}{TP + FP}, \quad (14)$$ где TP — количество истинно положительных исходов; FP — количество ложно положительных исходов	$Pr \in [0,1]$
Чувствительность, полнота (sensitivity, recall)	$$Se = \frac{TP}{TP + FN}, \quad (15)$$ где TP — количество истинно положительных исходов; FN — количество ложно отрицательных исходов	$Se \in [0,1]$
Избирательность (specificity)	$$Sp = \frac{TN}{TN + FP}, \quad (16)$$ где TN — количество истинно отрицательных исходов; FP — количество ложно положительных исходов	$Sp \in [0,1]$
F-мера	$$F = \frac{2 \cdot Pr \cdot Re}{Pr + Re}, \quad (17)$$ $$Pr = \frac{TP}{TP + FP},$$ $$Re = \frac{TP}{TP + FN},$$ где TP — количество истинно положительных исходов; FP — количество ложно положительных исходов; FN — количество ложно отрицательных исходов	$F \in [0,1]$

Продолжение таблицы 5

Наименование метрики	Формула	Измеряемый диапазон
Площадь под кривой ROC	$$AUCROC = \int_0^1 Se\,\partial F,$$ $$F = 1 - Sp = \frac{FP}{TN + FP},$$ (18) $$Se = \frac{TP}{TP + FN},$$ где *ROC*-кривая — график зависимости чувствительности от избирательности; *Se* — чувствительность; *Sp* — избирательность; *TP* — количество истинно положительных исходов; *TN* — количество истинно отрицательных исходов; *FP* — количество ложно положительных исходов; *FN* — количество ложно отрицательных исходов	$AUCROC \in [0,1]$
Площадь под кривой PRC	$$AUCPRC = \int_0^1 Pr\,\partial,$$ $$Pr = \frac{TP}{TP + FP},$$ (19) $$Se = \frac{TP}{TP + FN},$$ где *PRC*-кривая — график зависимости точности от чувствительности (полноты); *Pr* — точность; *Se* — чувствительность (полнота); *TP* — количество истинно положительных исходов; *FP* — количество ложно положительных исходов; *FN* — количество ложно отрицательных исходов	$AUCPRC \in [0,1]$
В задачах ранжирования		
Приведенная суммарная эффективность	$$DCG = \sum_{i=1}^{N} \frac{2^{rel_i} - 1}{\log_2(i+1)},$$ $$IDCG = \sum_{i=1}^{\lvert REL_N \rvert} \frac{2^{rel_i} - 1}{\log_2(i+1)},$$ (20) $$NDCG = \frac{DCG}{IDCG},$$ где N — количество ранжированных элементов; rel_i — оценка релевантности i-го элемента в ранжированном списке; REL_N — список релевантных элементов, упорядоченных по степени релевантности	$NDCG \in [0,1]$

Окончание таблицы 5

Наименование метрики	Формула	Измеряемый диапазон
В задачах восстановления (синтеза и реконструкции) изображений		
Пиковое отношение сигнал/шум	$$PSNR = \frac{10\lg\left(E_{max}^2\right)}{MSE},$$ $$MSE = \frac{1}{Kwh}\sum_{c=1}^{K}\sum_{x=0}^{w-1}\sum_{y=0}^{h-1}\left(I(x,y,c) - \hat{I}(x,y,c)\right), \quad (21)$$ $$E_{max} = 2^B - 1,$$ где B — разрядность (глубина квантования); E_{max} — максимальное значение яркости (интенсивности сигнала); I — выходные данные СИИ; $\hat{I}$ — эталонные (референсные) выходные данные; K — количество каналов (например, для монохромных изображений $K = 1$, в модели RGB $K = 3$)	—
Индекс структурного сходства	$$SSIM = C_l\left(I,\hat{I}\right)^{\alpha} \cdot C_c\left(I,\hat{I}\right)^{\beta} \cdot C_s\left(I,\hat{I}\right)^{\gamma},$$ $$\begin{cases} C_l\left(I,\hat{I}\right) = \dfrac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1} \\ C_c\left(I,\hat{I}\right) = \dfrac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2}, \\ C_s\left(I,\hat{I}\right) = \dfrac{cov\left(I,\hat{I}\right) + C_3}{\sigma_I\sigma_{\hat{I}} + C_3} \end{cases} \quad (22)$$ где I — выходные данные СИИ; $\hat{I}$ — эталонные (референсные) выходные данные; α, β, γ — весовые коэффициенты для каждого показателя, по умолчанию принимаются равными 1; C_1, C_2, C_3 — константы, которые следует вводить для предотвращения деления на ноль; μ_I и σ_I — среднее и среднеквадратическое отклонение выходных данных СИИ соответственно; $\mu_{\hat{I}}$ и $\sigma_{\hat{I}}$ — среднее и среднеквадратическое отклонение эталонных выходных данных соответственно	$SSIM \in [-1, +1]$

8.2.4 Метрики функциональной пригодности (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 (см. таблицу 6).

Таблица 6 — Примеры метрик оценки функциональной целесообразности

Наименование метрики	Формула	Измеряемый диапазон
Степень автоматизации	$$M_4 = \frac{A}{B}, \quad (23)$$ где A — количество шагов, выполняемых СИИ без привлечения пользователя, при реализации конкретной процедуры; B — общее количество шагов при выполнении заданной процедуры	$M_4 \in [0,1]$

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

8.3 Уровень производительности (performance efficiency) СИИ. Метрики

8.3.1 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.

8.3.2 В протокол тестирования следует включить параметры, влияющие на показатели качества: параметры ЦПУ, объем памяти, включая объем хранения (общий объем для хранения моделей ИИ), сетевой трафик и прочее.

8.3.3 Для оценки уровня производительности СИИ рекомендуется применять следующий набор субхарактеристик: характер изменения во времени (time behaviour), характер изменения (использования) ресурсов (resource utilization) и производительные возможности (capacity).

8.3.4 Метрики характер изменения во времени (time behaviour) используют для оценки степени соответствия требованиям временных ресурсов, затрачиваемых пользователем в целях обеспечения точной и полной реализации конкретных задач (см. таблицу 7).

Таблица 7 — Примеры метрик оценки временной эффективности

Наименование метрики	Формула	Измеряемый диапазон
Отклонение времени отклика	$$M_5 = \frac{\left\lvert \sum_{i=1}^{N} T_i - NT_н \right\rvert}{NT_н}, \quad (24)$$ где T_i — время отклика i-о измерения; N — количество измерений; $T_н$ — допустимое время отклика, представленного в технической документации	$M_5 > 0$
Производительность	$$M_6 = \frac{A}{T}, \quad (25)$$ где A — количество однотипных задач, выполненных СИИ за время T; T — время	$M_6 > 0$

8.3.5 Метрики характера изменения (использования) ресурсов (resource utilization) используют для оценки степени удовлетворения требований по потреблению объемов и продолжительности использования ресурсов СИИ при выполнении ее функций.

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

8.4 Способность СИИ к взаимодействию (compatibility). Метрики

8.4.1 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.

8.4.2 Для оценки совместимости способности к взаимодействию СИИ рекомендуется применять следующий набор субхарактеристик: соответствие (co-existence), функциональная совместимость (interoperability), контролируемость (controllability).

8.4.3 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.

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

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

8.5 Мобильность (portability) СИИ. Метрики

8.5.1 Для оценки мобильности СИИ рекомендуется применять следующий набор субхарактеристик: адаптируемость (adaptability), простота внедрения (installability), (взаимо)заменяемость (replacebility).

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

8.5.3 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. Метрики

8.6.1 Для оценки практичности СИИ рекомендуется применять следующий набор субхарактеристик: понятность (explainability), изучаемость (learnability), простота использования (operability), защищенность от ошибки пользователя (user error protection), эстетика пользовательского интерфейса (user interface aesthetics), доступность (accessibility), взаимодействие (collaborability).

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, несанкционированному 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 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 при оценке понятности для пользователя результатов работы СИИ.

8.7 Сопровождаемость (maintainability) СИИ. Метрики

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

8.7.2 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.

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

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

8.7.5 Метрики настраиваемости (evolution) используются для определения степени надежности мониторинга СИИ дрейфа данных, вычисления веса важности и правильности принятия меры для перезапуска процесса обучения или перестройки модели.

8.8 Надежность (reliability) СИИ. Метрики

8.8.1 Для оценки надежности СИИ рекомендуется применять следующий набор субхарактеристик 2-го рода: стабильность (maturity), устойчивость к ошибке (отказоустойчивость) (fault tolerance), восстанавливаемость (recoverability), робастность (robustness).

8.8.2 Метрики стабильности (maturity) используются для оценки независимости СИИ от отказов, существующих в самой СИИ (см. таблицу 8).

Таблица 8 — Примеры метрик стабильности

Наименование метрики	Формула	Измеряемый диапазон
Плотность отказов по отношению к тестовым наборам	$$M_7 = \frac{O}{T},$$ (26) где O — количество обнаруженных отказов СИИ; T — количество выполненных тестовых наборов	$M_7 \geq 0$
Устранение ошибок	$$M_8 = \frac{A}{B},$$ (27) где A — количество исправленных ошибок; B — общее количество реально обнаруженных (или спрогнозированных) ошибок	$M_8 \in [0,1]$
Тестовое покрытие	$$M_9 = \frac{A}{B},$$ (28) где A — 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	$M_9 \in [0,1]$
Коэффициент аварийных отказов	$$M_{10} = 1 - \frac{A}{B},$$ (29) где A — количество аварийных отказов СИИ; B — количество отказов СИИ	$M_{10} \in [0,1]$

8.8.3 Метрики устойчивости к ошибке (отказоустойчивости) (fault tolerance) используются для оценки способности СИИ поддерживать определенный уровень производительности в случаях возникновения ошибок и отказов оборудования при эксплуатации.

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

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

9 Требования к набору входных и выходных данных для тестирования СИИ

9.1 Для целей тестирования в процессе жизненного цикла СИИ применяются следующие наборы данных (см. таблицу 9):

- базовый демонстрационный набор данных — образцовый аннотированный набор данных, сформированный в соответствии с нормативно-техническим документом, устанавливающим унифицированные требования к проведению тестирования (испытаний) СИИ определенного типа;
- дополнительный демонстрационный набор данных — дополнительный аннотированный набор данных, предоставляемый заказчиком СИИ при формировании уточненных требований к системе с учетом 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 и обработки данных, позволяющие повысить качество создаваемой СИИ, а также технологии обогащения данных;
- тестовый набор данных — набор данных, формируемый в органах по оценке соответствия на основе демонстрационного набора данных и необходимый для проведения сертификации (регистрации), тестирования (испытаний) или аттестации СИИ.

Таблица 9 — Типы наборов данных

Тип набора данных	Назначение	Способ формирования
Базовый демонстрационный	Определение минимальных требований к функциональным характеристикам СИИ	В соответствии с нормативно-техническим документом (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 уточненных требований к СИИ с учетом 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 установленным требованиям	На основе демонстрационного набора данных в соответствии с требованиями сертификации (регистрации), тестирования (испытаний) или аттестации СИИ

9.2 При формировании тестового набора данных в целях обеспечения качества данных по [1] необходимо учитывать следующие требования к тестовым наборам данных:

- представительность: отклонение оценок функциональных характеристик СИИ, полученных при проведении тестирования (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 определенную, допустимо малую величину. К критериям представительности тестового набора данных относят достаточное совпадение статистических характеристик существенных условий эксплуатации для тестового набора данных и в предусмотренных условиях эксплуатации;
- безызбыточность: при формировании тестового набора данных необходимо стремиться к тому, чтобы его объем был минимально достаточным для выполнения требования представительности. Исключение составляют СИИ с измерительными функциями, например по ГОСТ 58450 и ГОСТ Р 8.673, для которых избыточность тестового набора данных требуется для повышения точности проводимых измерений;

- 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;
- требования по расчету размеров различных наборов входных и выходных данных;
- требования к характеристикам различных наборов входных и выходных данных.

Библиография

[1] ИСО/МЭК 25012:2008 Программная инженерия. Требования и оценка качества программного продукта (SQuaRE). Модель качества данных (ISO/IEC 25012:2008 Software engineering — Software product quality requirements and evaluation (SQuaRE) — Data quality model)

УДК 615.841:006.354 ОКС 11.040.01

Ключевые слова: системы искусственного интеллекта, контроль качества, метрики

Редактор *Н.А. Аргунова*
Технический редактор *И.Е. Черепкова*
Корректор *О.В. Лазарева*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 29.11.2021. Подписано в печать 29.12.2021. Формат 60×84⅛. 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 «РСТ»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.

gost.e-ecolog.ru gost.e-ecolog.ru

ОКС 11.040.01

Поправка к ГОСТ Р 59898—2021 Оценка качества систем искусственного интеллекта. Общие положения

В каком месте	Напечатано	Должно быть
Библиографические данные	ОКС 11.040.01	ОКС 35.020

(ИУС № 9 2025 г.)

gost.e-ecolog.ru