ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 59921.11— 2025

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ МЕДИЦИНЕ

Наборы данных для тестирования алгоритмов. Методы контроля набора данных на универсальность и структурированность

Издание официальное

Москва
Российский институт стандартизации
2025

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Государственным бюджетным учреждением здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ»)

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 28 октября 2025 г. № 1294-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 «Институт стандартизации», 2025

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 1
4 Общие положения 5
5 Процессы создания и использования наборов данных для тестирования систем искусственного интеллекта 6
6 Методы контроля наборов данных на универсальность и структурированность 8
6.1 Методы контроля наборов данных для тестирования систем искусственного интеллекта на структурированность 8
6.2 Методы контроля наборов данных для тестирования систем искусственного интеллекта на универсальность 10
7 Фрагменты тестовых наборов данных (демонстрационные тестовые наборы данных) 12
Приложение А (справочное) Пример структуры технического задания 13
Приложение Б (справочное) Пример описания версий набора данных 14
Приложение В (обязательное) Перечень критериев и порядок численной оценки процессов создания и использования наборов данных 15
Приложение Г (справочное) Пример структуры реестра наборов данных 16
Приложение Д (рекомендуемое) Чек-лист и балльная оценка структурированности набора данных . 17
Приложение Е (рекомендуемое) Чек-лист и балльная оценка универсальности набора данных 18
Библиография 19

Введение

Системы искусственного интеллекта в клинической медицине позволяют повысить качество и доступность медицинских услуг, а также оптимизируют ресурсы системы здравоохранения за счет автоматизации рутинных задач и анализа больших объемов медицинской информации с последующим предоставлением результатов этого анализа. Применение систем искусственного интеллекта позволяет повысить качество диагностического процесса за счет поддержки принятия врачебных решений, однако при этом с целью минимизации рисков, связанных с безопасностью пациентов, требуется проведение ряда тестирований, которые в свою очередь основываются на использовании репрезентативных, релевантных наборов данных. Настоящий стандарт устанавливает общие положения в процессах создания и использования таких наборов данных с целью проведения тестирования систем искусственного интеллекта, а также обеспечения доступности таких наборов данных для развития систем искусственного интеллекта.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КЛИНИЧЕСКОЙ МЕДИЦИНЕ

Наборы данных для тестирования алгоритмов. Методы контроля набора данных на универсальность и структурированность

Artificial intelligence in clinical medicine. Datasets for algorithm testing. Supervision over dataset structure and versatility

Дата введения — 2026—01—01

1 Область применения

Настоящий стандарт устанавливает общие требования к структуре процессов создания и методологии использования наборов данных для тестирования систем искусственного интеллекта (далее — СИИ, алгоритм). Положения настоящего стандарта могут быть адаптированы и использованы при формировании наборов данных для решения задач в предметной и смежной с предметной областях.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 8.736 Государственная система обеспечения единства измерений. Измерения прямые многократные. Методы обработки результатов измерений. Основные положения

ГОСТ Р 59921.5—2022 Системы искусственного интеллекта в клинической медицине. Часть 5. Требования к структуре и порядку применения набора данных для обучения и тестирования алгоритмов

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:

Издание официальное

3.1

аналитическая валидация: Подтверждение способности системы искусственного интеллекта точно, воспроизводимо и надежно генерировать предполагаемые результаты вычислений из входных данных.

Примечания

1 См. [1] [1)].

2 Аналитическая валидация является частным случаем валидации в соответствии с ГОСТ Р ИСО/МЭК 12207—2010, пункт 4.54.

[ГОСТ Р 59921.1—2022, пункт 3.1]

3.2

валидация: Подтверждение (на основе 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]

3.3

верификация: Подтверждение (на основе представления объективных свидетельств) того, что заданные требования полностью 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]

3.4 **внешняя валидация:** Валидация, осуществляемая в организации(ях), не принимавшей(их) участие в разработке систем искусственного интеллекта.

3.5 **внутренняя валидация:** Валидация, осуществляемая в организации(ях), принимавшей(их) участие в разработке систем искусственного интеллекта.

3.6 **запись набора данных:** Набор элементов набора данных, связанных единым уникальным идентификатором.

3.7 **инициирование набора данных:** Этап, запускающий процесс создания набора данных, на котором происходит определение перечня формальных признаков, необходимых и достаточных для его первичной идентификации.

3.8

искусственный интеллект: Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение, поиск и принятие решения без заранее заданного алгоритма) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

Примечание — Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных, анализу и синтезу решений.

[Адаптировано из ГОСТ Р 59277—2020, пункт 3.18]

[1)] IMDRF/SaMD WG/N41 — Software as a Medical Device (SaMD): Clinical Evaluation, 2017.

3.9

классификация: Способ и результат упорядочения, структуризации некоторого множества объектов, разделения его на определенные подмножества, выделения некоторых признаков объектов исходного множества как основания их структуризации по данному признаку.

Примечание — Такого рода признак называется основанием классификации.

[Адаптировано из ГОСТ Р 59277—2020, пункт 3.26]

3.10

клиническая валидация: Подтверждение способности системы искусственного интеллекта выдавать клинически значимые выходные данные, связанные с целевым использованием системы искусственного интеллекта в рамках установленного изготовителем функционального назначения.

Примечание — См. [1], пункт 7.0[1)].

[ГОСТ Р 59921.1—2022, пункт 3.10]

3.11

набор данных; НД: Совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта.

Примечание — Слово «необходимости» заменено на «необходимых» для унификации с другими терминами и определениями настоящего стандарта.

[Адаптировано из ГОСТ Р 59921.4—2021, пункт 3.5]

3.12

обеспечение качества: Совокупность систематических и планомерных действий, которые имеют целью обеспечить соответствие проведения исследования, сбора, регистрации и представления данных надлежащей клинической практике (GCP) и нормативным требованиям.

[ГОСТ Р 52379—2005, пункт 1.34]

3.13 **параметр набора данных:** Значение или категория, которая присваивается каждому элементу набора данных и обозначает его характеристику.

3.14 **планирование набора данных:** Этап жизненного цикла набора данных, на котором происходит формализация требований к входным и выходным данным, а также регламентирование процесса работы с данными.

3.15

проспективная разметка: Разметка данных в соответствии с поставленной целью формирования набора данных, предполагающая проведение дополнительных манипуляций с элементами.

Примечания

1 Проспективную разметку выполняют путем постановки метки начала и окончания события, меток обнаружения признаков, обозначений патологий и т. п.

2 Слово «сбор» заменено на «разметка и «а также» на «предполагающая» для унификации с другими терминами настоящего стандарта.

[Адаптировано из ГОСТ Р 59921.0—2022, пункт 3.3.10]

3.16 **публикация набора данных:** Этап жизненного цикла набора данных, заключающийся в размещении сформированного набора данных в хранилище.

[1)] IMDRF/SaMD WG/N41 — Software as a Medical Device (SaMD): Clinical Evaluation, 2017.

3.17

разметка [аннотация] данных: Этап обработки структурированных и неструктурированных данных, в процессе которого данным (в том числе текстовым документам, фото- и видеоизображениям) присваиваются идентификаторы, отражающие тип данных (классификация данных), и (или) осуществляется интерпретация данных для решения конкретной задачи, в том числе с использованием систем искусственного интеллекта.

[ГОСТ Р 59921.5—2022, пункт 3.19]

3.18 **регистрация набора данных:** Этап жизненного цикла набора данных, фиксирующий факт завершения работ по созданию набора данных в сопутствующей документации.

3.19 **реестр наборов данных:** Систематизированный перечень сведений обо всех наборах данных организации/подразделения, ведущийся уполномоченным сотрудником с целью упорядочивания деятельности по формированию и использованию наборов данных.

3.20 **релевантный набор данных:** Набор данных, достаточный для достижения установленной цели.

3.21 **репрезентативный набор данных:** Набор данных, характеристики выборки которого должны соответствовать характеристикам генеральной совокупности.

3.22

ретроспективная разметка: Разметка данных в соответствии с указанными метаданными, перечень которых выбирают в соответствии с поставленной целью формирования набора данных.

Примечания

1 Ретроспективная разметка не предполагает дополнительных манипуляций с элементами данных (например, постановка метки начала и окончания события, меток обнаружения признаков, обозначений патологий и т. п.).

2 Слово «сбор» заменено на «разметка» для унификации с другими терминами настоящего стандарта.

[Адаптировано из ГОСТ Р 59921.0—2022, пункт 3.3.9]

3.23

сбор данных: Процесс объединения данных, поступающих из одного или более источников, в целях их использования при обучении и тестировании систем искусственного интеллекта.

[ГОСТ Р 59921.5—2022, пункт 3.24]

3.24

система искусственного интеллекта: Программное обеспечение, в котором используются технологические решения искусственного интеллекта.

[ГОСТ Р 59921.6—2021, пункт 3.10]

3.25 **система версионности набора данных:** Система, фиксирующая изменения в наборе данных путем сохранения различных его версий с целью обеспечения воспроизводимости процесса создания набора данных.

3.26

система менеджмента качества систем искусственного интеллекта: Организационная структура, функции, процедуры, процессы и ресурсы, необходимые для скоординированной деятельности по руководству и управлению производителем системы искусственного интеллекта применительно к качеству.

[ГОСТ Р 59921.5—2022, пункт 3.26]

3.27 **смена версии набора данных:** Этап жизненного цикла набора данных, на котором запускается процесс внесения изменений в уже готовый набор данных.

3.28 **структурированность процессов создания и использования набора данных:** Разделение жизненного цикла набора данных на регламентированные этапы.

3.29 **структурированность набора данных:** Организация набора данных по определенным правилам и с использованием определенных методов, представленных в техническом задании на подготовку набора данных.

3.30

тестирование: Деятельность, в которой система или компонент выполняется в определенных условиях, результаты наблюдают или фиксируют, а также проводят оценку какого-либо аспекта системы или компонента.

[ГОСТ Р 70462.1—2022, пункт 3.7]

3.31 **техническое задание на набор данных;** ТЗ: Исходный документ для проведения работы, устанавливающий требования к создаваемому набору данных и технической документации на него, а также требования к объему, срокам проведения работы и форме представления результатов.

3.32 **универсальность набора данных для тестирования систем искусственного интеллекта:** Возможность использования набора данных для решения более чем одной задачи.

3.33 **универсальность процессов создания и использования набора данных:** Применимость этапов жизненного цикла для создания набора данных в различных задачах.

3.34 **формирование:** Этап жизненного цикла наборов данных, заключающийся непосредственно в его создании в соответствии с регламентирующими документами этапа планирования.

3.35 **элемент набора данных:** Минимальная структурная единица набора данных, определяемая как значение конкретного параметра набора данных для соответствующей записи набора данных.

4 Общие положения

4.1 Обеспечение качества и безопасности СИИ в клинической медицине требует проведения их валидации, включая тестирование с использованием специализированных репрезентативных, релевантных медицинских НД. Процессы создания и использования таких НД приведены в разделе 5.

4.2 Валидацию СИИ по месту проведения подразделяют на следующие виды:

- внутреннюю валидацию (зависимую);
- внешнюю валидацию (независимую).

4.3 Валидацию СИИ по оцениваемым параметрам подразделяют на следующие виды:

- клиническую валидацию;
- аналитическую валидацию.

4.4 Валидацию в зависимости от этапа жизненного цикла СИИ подразделяют на следующие виды:

- валидация на этапе разработки;
- валидация на этапе внедрения;
- валидация на этапе эксплуатации (мониторинг).

4.5 Для совершенствования системы менеджмента качества в сфере тестирования СИИ (в том числе с использованием методов автоматизации) необходимы структурированность и универсальность НД (принципы универсальности и структурированности НД приведены в разделе 6).

4.6 Структурированность процессов создания и использования НД основывается на жизненном цикле НД, приведенном на рисунке 1.

4.7 Универсальность процессов создания определяется способностью НД соответствовать этапам жизненного цикла.

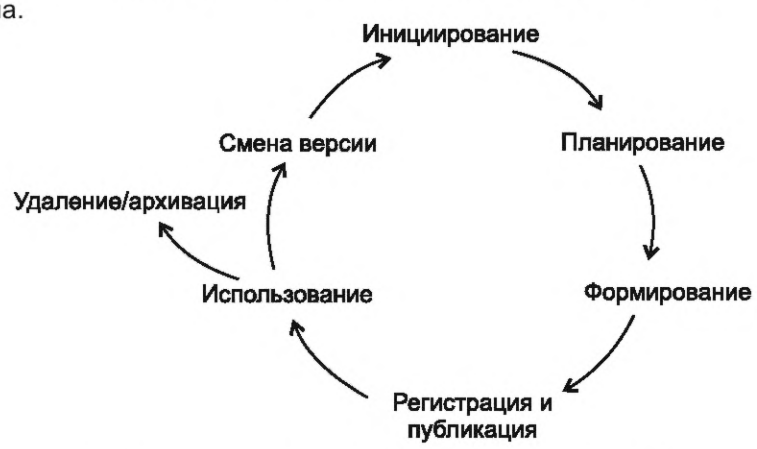


Рисунок 1 — Этапы жизненного цикла набора данных

5 Процессы создания и использования наборов данных для тестирования систем искусственного интеллекта

5.1 Инициирование

5.1.1 На этапе инициирования НД могут быть определены:

- наименование НД;
- уникальный идентификатор НД;
- назначение НД;
- цель создания НД;
- разработчик НД;
- перечень участников (в т. ч. ответственные исполнители) процесса создания НД и др.

5.1.2 В соответствии с классификацией, приведенной в 4.3, на этапе инициирования следует учитывать клинические аспекты применения и технические характеристики СИИ, тестируемой с использованием НД. Данные требования являются основой для разработки комплекта технической документации на НД на этапе планирования.

5.2 Планирование

5.2.1 На этапе планирования проводят формализацию требований к НД. Они могут быть представлены в виде комплекта технической документации, включающего ТЗ на создание НД и приложений к нему. Одна из возможных структур ТЗ приведена в приложении А. При этом важной составляющей ТЗ является описание структуры планируемого НД, которая в дальнейшем определит структурированность готового НД (см. 6.1).

5.2.2 Дополнительно на этапе планирования НД могут быть определены:

- методическое обеспечение процесса подготовки НД в формате обучающих материалов;
- перечень инструментов и сроки промежуточного контроля процесса подготовки НД;
- требования к объему НД (определяются исследуемой при тестировании СИИ метрикой и доверительным интервалом, статистической значимостью, мощностью статистического теста и балансом классов) и др.

Примечания

1 При работе с разнородными данными (например, медицинскими данными) следует руководствоваться стандартными подходами к оценке необходимого и достаточного объема выборки, приведенными в ГОСТ Р 59921.5.

2 При работе со стандартными образцами, например, тест-объектами, следует руководствоваться основными положениями ГОСТ Р 8.736.

5.3 Формирование

5.3.1 На этапе формирования НД могут быть проведены:

- сбор данных;
- деидентификация данных;
- фильтрация данных;
- разметка данных;
- верификация результатов разметки;
- структурирование данных.

5.3.1.1 Различают следующие типы сбора данных:

- проспективный, заключающийся в выполнении клинических и/или диагностических исследований по заданным параметрам;
- ретроспективный, заключающийся в поиске, фильтрации по заданным параметрам (например, фильтрация текстовых протоколов лучевых исследований по ключевым словам с помощью алгоритмов обработки естественного языка) и выгрузке ранее проведенных исследований (например, из медицинской информационной системы).

5.3.1.2 Деидентификация данных проводится в соответствии с [1], а также с учетом ГОСТ Р 59921.5—2022 (подраздел 5.7).

5.3.1.3 Фильтрация данных может осуществляться как при сборе данных (см. 5.3.1.1), так и представлять собой самостоятельный процесс.

Пример — Проверка корректности заполнения DICOM-тегов (Digital Imaging and Communications in Medicine) в лучевых исследованиях с помощью инструментов автоматического анализа.

5.3.1.4 Различают следующие типы разметки:

- проспективная;
- ретроспективная.

Для проспективной разметки требуются дополнительные манипуляции с участием медицинского персонала, требования к квалификации которого приведены в ГОСТ Р 59921.5—2022 (подраздел 7.2). Количество персонала, его квалификация и дизайн процесса разметки зависят от конкретной задачи и метода верификации. Как правило, минимальное рекомендуемое количество разметчиков — два медицинских специалиста.

Ретроспективная разметка проводится в соответствии с имеющимися метаданными [например, извлечение из медицинской карты клинической информации (количество родов, возраст наступления менопаузы, код МКБ-10 [2])].

5.3.1.5 Разметку данных верифицируют исходя из клинической задачи с учетом «золотого стандарта» выявления патологии/признака. Верификацию результатов разметки проводят с помощью следующих методов:

- согласно тексту описания исследования (например, заключения ультразвукового исследования);
- путем экспертного пересмотра (например, пересмотр врачом-разметчиком и экспертом);
- путем исследования того же вида в динамике (например, признаки кровоизлияния в заключении компьютерной томографии в динамике);
- путем исследования другой модальности (например, для рентгенографического исследования — верификация данных на компьютерной томограмме);
- путем лабораторного исследования (например, данные патоморфологического исследования для верификации злокачественных новообразований на лучевом исследовании);
- согласно клиническому диагнозу (например, установленный диагноз U07.1 по данным медицинской карты).

5.3.1.6 Структурирование данных включает:

- проверку результатов разметки на полноту, отсутствие дубликатов, ошибок, достаточность и соответствие ТЗ;
- объединение результатов разметки;
- формирование итоговых файлов разметки;
- формирование файлов данных и разметки путем создания репозитория со структурированными файлами разметки, исследований и другими дополнительными данными (например, маски, в случае сегментации изображений);
- создание сопроводительного текстового файла (readme), содержащего краткую структурированную информацию о НД (структурирование метаданных).

Примечание — Readme-файл хранится вместе с файлами НД с целью предоставления информации конечному пользователю о НД.

5.4 Регистрация

Этап регистрации НД предполагает размещение сформированного НД в хранилище (архив), а также фиксацию факта завершения работ в сопутствующей документации. На данном этапе проводится проверка готового НД на соответствие требованиям ТЗ на НД. Организации — изготовителю НД следует вести специальный реестр, содержащий структурированную и стандартизованную информацию о НД. Информацию в реестре следует синхронизировать с этапами жизненного цикла НД. Ведение реестра обеспечивает:

- контроль соответствия требованиям к НД;
- управление процессами создания и использования НД;
- централизованный доступ к данным;
- автоматизацию процессов, связанных с созданием и использованием НД.

5.5 Публикация

На этапе публикации НД его размещают для последующего использования в локальном хранилище с предоставлением доступа ограниченному кругу лиц в соответствии с требованиями действующего

законодательства и локальных нормативных актов или в открытом доступе (например, в специализированных библиотеках НД) с учетом соблюдения требований [3].

5.6 Использование

Этап использования заключается в непосредственном применении НД в соответствии с поставленной целью. Возможно использование одного и того же НД (или его части) для разных задач с целью оптимизации использования ресурсов, однако во избежание различных конфликтов интересов или ошибок (например, применение данных для тестирования, если они использовались для обучения СИИ) необходимо фиксировать эту информацию в реестре НД или специальных журналах.

5.7 Удаление и архивация

Неактуальные НД могут быть заархивированы или удалены. Удаление неактуальных НД не рекомендуется, так как в дальнейшем может потребоваться восстановление данных.

5.8 Смена версии

Для обеспечения надежности и воспроизводимости процессов создания и использования НД при внесении в них изменений (все изменения должны быть задокументированы и приложены к НД) следует использовать систему версионности. Пример описания версий НД приведен в приложении Б.

5.9 Оценка структурированности и универсальности процессов создания и использования наборов данных для тестирования систем искусственного интеллекта

Перечень критериев и порядок их численной оценки должен соответствовать приведенному в приложении В.

6 Методы контроля наборов данных на универсальность и структурированность

6.1 Методы контроля наборов данных для тестирования систем искусственного интеллекта на структурированность

6.1.1 Для клинических данных характерна гетерогенность, зависимость от контекста и в ряде случаев неполнота (незавершенность). В целях статистического и/или автоматизированного анализа, в том числе с использованием СИИ, исходные клинические данные необходимо трансформировать в структурированную форму. Структурированность НД для тестирования СИИ подразумевает структурированность как элементов НД по 6.1.2—6.1.6, так и метаданных по 6.1.6. Структурировать элементы НД необходимо в соответствии:

- с наименованием целевых признаков НД и их взаимосвязями (в соответствии с принятыми медицинскими справочниками номенклатур и стандартами);
- видом (бинарная, мультиклассовая, регрессионная, сегментация, локализация и т. д.) и уровнем разметки (изображение, исследование, пациент и т. д.);
- формой представления разметки (логическая, численная, текст; единицы измерения, количество значащих цифр, диапазон значений и т. д.).

Для обеспечения структурированности элементов НД следует руководствоваться критериями, приведенными в 6.1.2—6.1.6.

6.1.2 НД должен удовлетворять критерию системности — каждый элемент НД должен:

- обладать четко определенным набором связей с другими элементами записи;
- не должен быть изолированным;
- должен быть организован в соответствии с едиными правилами, описанными в ТЗ.

В НД, удовлетворяющем критерию системности, поиск любого элемента возможен с использованием связей внутри системы.

Пример — По индивидуальному номеру пациента возможно отследить все его исследования в системе.

Коэффициент системности НД $k_{сист}$ вычисляют по формуле

$$k_{\text{сист}} = \frac{n_{\text{связ}}}{n_{\text{элем}} - 1}, \quad (1)$$

где $n_{\text{связ}}$ — количество связей в НД;

$n_{\text{элем}}$ — количество элементов в записях НД.

Так как минимальное число условных простейших связей в системе (когда каждый элемент системы должен быть связан только с одним элементом и только одной связью) на единицу меньше числа элементов в ней, то значение отношения $\frac{n_{\text{связ}}}{n_{\text{элем}} - 1}$ будет менее единицы при наличии изолированных элементов в системе, равно единице в случае наличия минимума простейших связей и более единицы во всех остальных случаях. Соблюдение критерия системности в этом случае определяется по значению коэффициента системности следующим образом:

- $k_{\text{сист}} > 1$ — полное соответствие;
- $k_{\text{сист}} = 1$ — частичное соответствие;
- $k_{\text{сист}} < 1$ несоответствие.

Примечание — Для наборов медицинских данных не требуется полное соответствие критерию системности, так как поиск элемента системы по любому другому элементу системы избыточен. Необходимым и достаточным является наличие иерархических связей, позволяющих однозначно определить любой элемент записи НД по ее (записи) идентификатору.

6.1.3 НД должен удовлетворять критерию уникальности записей.

Пример — Идентификационный номер диагностического исследования в НД должен быть ассоциирован с одним конкретным пациентом.

Коэффициент уникальности записей НД $k_{\text{ун.зап}}$ вычисляют по формуле

$$k_{\text{ун.зап}} = \frac{n_{\text{идент}}}{n_{\text{зап}}}, \quad (2)$$

где $n_{\text{идент}}$ — количество уникальных идентификаторов в НД;

$n_{\text{зап}}$ — количество идентифицированных записей в НД.

Значение $k_{\text{ун.зап}}$ будет равно единице, если каждой индексированной записи НД соответствует свой уникальный идентификатор. Значение $k_{\text{ун.зап}}$ будет менее единицы, если у нескольких записей будет одинаковый идентификатор. Соблюдение критерия уникальности записей в этом случае определяется по значению коэффициента уникальности записей следующим образом:

$k_{\text{ун.зап}} = 1$ — соответствует;

$k_{\text{ун.зап}} < 1$ — не соответствует;

6.1.4 НД должен удовлетворять критерию однозначности формата представления данных, то есть все элементы НД должны:

- быть представлены строго в соответствии с заранее определенными форматами представления данных;
- содержать значения, которые однозначно соответствуют допустимым величинам или диапазонам значений.

Для проверки соответствия критерию допустимо применять различные логические правила.

Пример — Проверка записи даты на соответствие формату «ДД-ММ-ГГГГ».

Коэффициент однозначности формата НД $k_{\text{од.ф}}$ вычисляют по формуле

$$k_{\text{од.ф}} = \frac{n_{\text{с.ф}}}{n_{\text{общ}}} + \frac{n_{\text{с.д}}}{n_{\text{общ}}}, \quad (3)$$

где $n_{\text{с.ф}}$ — количество записей НД, формат всех элементов которых соответствует требованиям (согласно ТЗ);

$n_{\text{с.д}}$ — количество записей НД, значения всех элементов которых соответствуют требуемому диапазону (согласно ТЗ);

$n_{\text{общ}}$ — общее количество записей в НД.

Значение $k_{\text{од.ф}}$ будет равно двум, если все записи НД, содержащие данные, будут представлены в требуемом формате и при этом их значения будут соответствовать заданному диапазону. Значение $k_{\text{од.ф}}$ будет менее двух, если в НД будут записи, для которых не выполняется хотя бы одно из условий соответствия формату или диапазону. Соблюдение критерия однозначности формата в этом случае определяется по значению коэффициента однозначности формата следующим образом:

- $k_{\text{од.ф}} = 2$ — соответствует;
- $k_{\text{од.ф}} < 2$ — не соответствует.

6.1.5 НД должен удовлетворять критерию соответствия поставленной задаче — структура НД для тестирования должна обеспечивать получение объективного, измеримого и однозначного результата, отвечающего поставленной задаче.

Для проверки соответствия критерию допустимо применять методы сопоставления.

Пример — Проверка соответствия НД задаче, изложенной в ТЗ.

Коэффициент соответствия поставленной задаче $k_{\text{с.з}}$ вычисляют по формуле

$$k_{\text{с.з}} = \frac{n_{\text{п.с.з}}}{n_{\text{общ}}}, \quad (4)$$

где $n_{\text{п.с.з}}$ — количество параметров НД, которые соответствуют поставленной задаче (согласно ТЗ);

$n_{\text{общ}}$ — общее количество параметров задачи (согласно ТЗ).

Значение $k_{\text{с.з}}$ будет равно единице, если всем параметрам поставленной задачи в однозначное соответствие можно привести параметры НД. Значение $k_{\text{с.з.}}$ будет менее единицы, если НД будет содержать параметры не полноценно отвечающие параметрам поставленной задачи. Соблюдение критерия соответствия поставленной задаче в этом случае определяется по значению коэффициента соответствия поставленной задаче следующим образом:

- $k_{\text{с.з}} = 1$ — полностью соответствует;
- $k_{\text{с.з}}$ в диапазоне от 0,5 до 1,0 — частично соответствует;
- $k_{\text{с.з}} < 0,5$ — не соответствует.

6.1.6 Структурировать метаданные НД необходимо в соответствии:

- с назначением;
- условиями доступа;
- клиническими параметрами;
- популяционными параметрами;
- техническими параметрами;
- параметрами разметки;
- данными об использовании.

Структурирование метаданных обеспечивается путем:

- ведения реестра НД по 5.4 (состав полей реестра определяется спецификой организации, в которой создаются и/или используются НД, и решаемыми с помощью НД задачами (чем шире круг задач, тем больше полей требуется в реестре). Пример состава реестра приведен в приложении Г, а также в [4];
- создания структурированного readme-файла по 5.3.1.6.

6.1.7 Для оценки структурированности НД для тестирования СИИ рекомендуется использовать чек-лист и балльную оценку, приведенные в приложении Д.

6.1.8 Периодичность контроля структурированности НД соответствует периодичности внесения изменений в НД. После каждого внесения изменений необходимо проводить контроль качества с целью соответствия новой версии НД критериям структурированности. Дополнительный контроль избыточен, так как НД не изменяет свои характеристики с течением времени.

6.2 Методы контроля наборов данных для тестирования систем искусственного интеллекта на универсальность

6.2.1 Под универсальностью НД для тестирования СИИ понимается возможность использования НД для решения более чем одной задачи. Для обеспечения универсальности НД следует контролировать соблюдение пригодности НД для решения нескольких различных задач.

Пример — Универсальным можно считать НД компьютерной томографии, используемый для тестирования двух СИИ, нацеленных на детектирование различных органов в области сканирования.

Для обеспечения универсальности НД следует руководствоваться критериями, приведенными в 6.2.2, 6.2.3.

6.2.2 НД должен удовлетворять критерию соответствия требованиям к тестируемой СИИ — каждый элемент НД должен:

- соответствовать формату и количеству входных данных, которые обрабатывает СИИ;
- соответствовать формату и количеству выходных данных, которые получаются в результате работы СИИ.

Пример — В требованиях к СИИ указано, что она должна обрабатывать данные рентгенографии в формате DICOM, при этом результат работы СИИ должен выводиться в виде бинарной классификации 1 — «наличие патологии», 0 — «отсутствие патологии», соответственно НД также должен содержать данные рентгенографии в формате DICOM и соответствующую бинарную классификацию.

Коэффициент соответствия НД требованиям к тестируемой СИИ $k_{с.т}$ вычисляют по формуле

$$k_{с.т} = \frac{n_{п.с.т}}{n_{общ}}, \quad (5)$$

где $n_{п.с.т}$ — количество параметров НД, которые соответствуют требованиям к СИИ;

$n_{общ}$ — общее количество параметров НД.

Значение $k_{с.т}$ будет равно единице, если всем параметрам к тестируемой СИИ, оцениваемым с помощью конкретного НД, в однозначное соответствие можно привести параметры этого НД. Значение $k_{с.т}$ будет менее единицы, если НД будет содержать параметры не полноценно отвечающие параметрам поставленной задачи. Соблюдение критерия соответствия поставленной задаче в этом случае определяется по значению коэффициента соответствия поставленной задаче следующим образом:

- $k_{с.т} = 1$ — полностью соответствует;
- $k_{с.т}$ в диапазоне от 0,5 до 1,0 — частично соответствует;
- $k_{с.т} < 0,5$ — не соответствует.

Пример — Для задачи оценки диагностической точности выявления признаков, коррелирующих с наличием злокачественных новообразований в легких определены требования к параметрам СИИ: вывести значение наличия или отсутствия признака, определить диаметр каждого узла, объем узла, классифицировать его по RADS. Для $k_{с.т} = 1$ в НД для тестирования необходимо наличие тех же параметров.

6.2.3 Универсальность НД обеспечивается в том числе соблюдением принципов FAIR (англ. Findable, Accessible, Interoperable, Re-usable — находимость, доступность, совместимость и переиспользование). Принципы FAIR предполагают наличие в НД описания следующих параметров:

- источника (происхождения) данных, в том числе места и даты создания, и соответствующих протоколов медицинских процедур;
- определения единиц представления;
- каждого информационного поля НД;
- терминов/онтологий, используемых при создании НД, которые должны быть удобочитаемыми, уникальными, интуитивно понятными, однозначно интерпретируемыми и привязанными к возможным вариациям в клинических данных;
- семантических связей и/или единиц представления (кодов, идентификаторов, ярлыков), используемых при создании НД;
- единых синтаксических правил создания глоссариев НД, описывающих принципы комбинирования отдельных терминов.

Пример — Синтаксические правила для данных лабораторной диагностики, утверждающие порядок следования данных для наименования аналитов (компонентов анализируемого вещества), соответствующих значений и единиц измерения (например, онтология SNOMED CT содержит термин «первичное злокачественное новообразование легкого (заболевание)», однако медицинские тексты могут содержать различные вариации и синонимы этого термина, такие как «ЗНО легкого», «карцинома легких» и т. д.).

6.2.4 Для оценки универсальности НД для тестирования СИИ рекомендуется использовать чеклист и балльную оценку, приведенные в приложении Е.

6.2.5 При сравнительной оценке универсальности НД рекомендуется использовать нормированный критерий универсальности НКУ, вычисляемый по формуле

$$НКУ = n/n_{max}, \quad (6)$$

где n — количество задач, решаемых с помощью исследуемого НД;

n_{max} — максимальное количество задач, решаемых с помощью НД, среди всех НД, для которых производится сравнительная оценка.

Пример — НД1 содержит бинарную классификацию компьютерно-томографических исследований органов грудной клетки с наличием и отсутствием признаков пневмоторакса и гидроторакса в табличном виде: номер исследования и его класс («норма» и «патология») для каждого из двух признаков. НД2 содержит бинарную классификацию компьютерно-томографических исследований органов грудной клетки с наличием и отсутствием признаков пневмоторакса и гидроторакса в табличном виде, дополненную результатами измерений: номер исследования, его класс («норма» и «патология») для каждого из двух признаков, объем поражения по каждому из признаков в кубических миллилитрах. Оба НД допускается считать универсальными, так как они могут быть использованы для решения как минимум двух задач: оценки наличия в исследовании пневмоторакса и гидроторакса. Однако НД2 допускается использовать для решения еще двух дополнительных задач: оценки объема поражения при наличии каждого из признаков. Следовательно, n_{max} следует принять равным 4, тогда НКУ(НД1) будет равен 2/4 = 0,5, а НКУ(НД2) = 4/4 = 1.

6.2.6 Периодичность контроля на универсальность соответствует периодичности внесения изменений в НД. После каждого внесения изменений необходимо проводить контроль качества с целью соответствия критериям универсальности новой версии НД. Дополнительный контроль избыточен, так как НД не изменяет свои характеристики с течением времени.

7 Фрагменты тестовых наборов данных (демонстрационные тестовые наборы данных)

7.1 Основной демонстрационный тестовый НД для тестирования алгоритмов искусственного интеллекта, содержащий результаты маммографических исследований с наличием и отсутствием признаков рака молочных желез, а также дополнительные клинические данные (код МКБ-10 [2], возраст наступления менопаузы, количество родов), приведен в [5].

Приложение А
(справочное)

Пример структуры технического задания

1 Общая информация
1.1 Наименование НД
1.2 Назначение НД
1.3 Сроки создания НД
1.4 Языки НД
1.5 Режим доступа к НД
1.6 Данные разработчика НД
1.7 Термины и определения
2 Входные данные
2.1 Источник получения данных
2.2 Порядок доступа к данным
2.3 Тип входных данных
2.4 Формат входных данных
2.5 Структура входных данных
2.6 Этические и юридические ограничения работы с данными
3 Обработка данных
3.1 Требования к результатам обработки данных
3.2 Число участников обработки данных
3.3 Требования к участникам процесса обработки данных
3.4 Этапы обработки данных
3.5 Требования к инструментам обработки данных
3.6 Перечень лиц, ответственных за выполнение каждого этапа
4 Выходные данные
4.1 Формат представления выходных данных
4.2 Методы валидации выходных данных
4.3 Объем и структура выходных данных
4.4 Порядок размещения и передачи выходных данных
5 Вариативная часть ТЗ (может содержать следующую информацию)
5.1 Версия НД
5.2 Ключевые слова НД
5.3 Классификация НД
5.4 Порядок и инструменты деперсонализации данных
5.5 Порядок формирования итогового решения при вовлечении нескольких участников процесса обработки данных
5.6 Дополнительная обработка данных и другое

Приложение Б
(справочное)

Пример описания версий набора данных

Б.1 При смене версии следует использовать трехзначные значения формата А.Б.В (указывают в названии/идентификаторе НД), где А — мажорная версия, Б — минорная версия, В — патч-версия.

Б.2 Мажорная версия увеличивается при изменении значимых параметров НД, связанных с клинической задачей, целью, принципами разметки и верификации данных.

Б.3 Минорная версия увеличивается при замене, добавлении и удалении единиц данных (изображений, текстовых или сигнальных данных и др.) без изменений значимых параметров НД (минорная версия устанавливается в значение 0 при выпуске новой мажорной версии).

Б.4 Патч-версия увеличивается при внесении корректировок в сопроводительную документацию, исправлении опечаток или ошибок в файлах разметки и верификации, при этом не меняется ни количество, ни качество входных данных (патч-версия устанавливается в значение 0 при выпуске новой минорной или мажорной версии).

Приложение В
(обязательное)

Перечень критериев и порядок численной оценки процессов создания и использования наборов данных

В.1 Наличие сопроводительной информации (регламент, ТЗ или иной вид документирования), описывающей процесс создания НД:

- представлена в полном объеме и структурирована — 2;
- имеется, но она недостаточная и/или не структурирована — 1;
- отсутствует — 0.

В.2 Соблюдение этапов инициирования, планирования, формирования и регистрации НД в процессе создания:

- представлены в полном объеме — 2;
- представлены частично — 1;
- отсутствуют — 0.

В.3 Использование системы смены версионности в случае внесения изменений:

- используется — 1;
- не используется — 0.

В.4 Наличие фиксации информации об использовании НД в специальных журналах, реестрах и т. д.:

- имеется — 1;
- не имеется — 0.

В.5 Хранение НД в специализированных хранилищах (локальных или публичных) с учетом обеспечения безопасности в соответствии с [3]:

- наличие специализированного хранилища и регламента доступа к данным — 1;
- отсутствие специального хранилища и/или нарушение мер безопасности — 0.

Общая оценка процессов создания и использования НД формируется как сумма баллов по каждому пункту В.1—В.5. Итоговый балл находится в диапазоне [0; 7]. Приемлемой считается оценка не менее пяти баллов, при этом недопустимо отсутствие баллов хотя бы по одному пункту.

Приложение Г
(справочное)

Пример структуры реестра наборов данных

Структура реестра НД включает:

- название и/или идентификатор НД;
- цель создания и назначение НД (задачи, решаемые с помощью НД);
- условия доступа;
- версию НД;
- идентификаторы и/или наименования исследований в соответствии с Федеральными справочниками;
- клиническую информацию в соответствии с Федеральными справочниками и международными стандартами медицинских номенклатур (анатомическую локализацию, целевые и сопутствующие патологии и признаки, в том числе в соответствии с международной классификацией болезней);
- источник данных (например, пациенты, тест-объекты, результаты синтетической генерации и др.);
- популяционные параметры [половозрастной состав НД, географию сбора (медицинские организации и/или регион), эпидемиологические данные, критерии включения/невключения пациента и др.];
- параметры разметки (тип и характер разметки, количество и вид измеряемых параметров, методы верификации, критерии отнесения к классам, количество специалистов и др.);
- технические параметры (объект и результат разметки, форматы представления данных, количество записей, объем НД, диагностические устройства, информацию об обезличивании и др.);
- данные об использовании (количество проведенных тестирований, виды тестирований, проекты, сотрудничества, публикации, информацию о регистрации НД как результата интеллектуальной деятельности, ссылки на хранение, в том числе в открытом доступе и др.);
- информацию об ответственных и процессе подготовки [организации(ях) — исполнителе(ях), заказчике, ответственном исполнителе, информацию о сроке и статусе выполнения работ, финансировании, авторах и др.] и т.д.

Приложение Д
(рекомендуемое)

Чек-лист и балльная оценка структурированности набора данных

Д.1 Соответствие критерию системности организации элементов НД:
- $k_{сист} \geq 1$ — соответствие — 1;
- $k_{сист} < 1$ — несоответствие — 0.

Д.2 Соответствие критерию уникальности записей НД:
- $k_{ун.зап} = 1$ — соответствует — 1;
- $k_{ун.зап} < 1$ — не соответствует — 0.

Д.3 Соответствие критерию однозначности формата представления данных:
- $k_{од.ф} = 2$ — соответствует — 1;
- $k_{од.ф} < 2$ — не соответствует — 0.

Д.4 Соответствие поставленной задаче, для которой создан НД:
- $k_{с.з} = 1$ — полное соответствие — 2;
- $k_{с.з}$ в диапазоне от 0,5 до 1,0 — частичное соответствие — 1;
- $k_{с.з} < 0,5$ — несоответствие — 0.

Д.5 Элементы НД отображаются следующим образом (с множественным выбором):
- наименования целевых признаков и их взаимосвязи сформированы в соответствии со стандартами (справочники, приказы и т. д.) — 1;
- параметры разметки (вид, уровень, форма и т. д.) определены и едины во всем НД — 1.

Д.6 Сопроводительная информация (метаданные) к НД:
- представлена в полном объеме и структурирована — 2;
- имеется, но она недостаточная и/или не структурирована — 1;
- отсутствует — 0.

Общая оценка структурированности НД формируется как сумма баллов по каждому пункту Д.1—Д.6. Итоговый балл находится в диапазоне [0; 9]. Приемлемой считается оценка не менее пяти баллов, при этом недопустимо отсутствие баллов хотя бы по одному пункту.

Приложение Е
(рекомендуемое)

Чек-лист и балльная оценка универсальности набора данных

Е.1 Количество возможных задач, решаемых с применением НД:
- более одной задачи — 1;
- одна задача — 0.

Е.2 Соответствие НД требованиям к тестируемой СИИ (выбор одного из списка):
- полное соответствие ($k_{С.Т}$ = 1) — 2;
- частичное соответствие ($k_{С.Т}$ в диапазоне от 0,5 до 1) — 1;
- несоответствие ($k_{С.Т}$ < 0,5) — 0.

Е.3 Соответствие принципам FAIR (с множественным выбором):
- находимость — 1;
- доступность — 1;
- совместимость — 1;
- переиспользование — 1.

Е.4 Соответствие используемых терминов и их семантических связей, используемых в НД и в сопроводительной информации, общепринятым стандартам (справочники, нормативные документы) (с множественным выбором):
- единообразие терминологии и номенклатуры в НД и сопроводительной документации — 1;
- отсутствие неописанных синонимов — 1;
- все термины в НД используются в соответствии с общепринятыми справочниками — 1;
- отсутствие некорректных логических и семантических связей между терминами — 1.

Е.5 Наличие сопроводительной информации, в которой указано описание (с множественным выбором):
- источника (происхождения) данных, в том числе места и даты создания, и соответствующих протоколов медицинских процедур — 1;
- определения единиц представления — 1;
- каждого информационного поля НД — 1.

Общая оценка универсальности НД формируется как сумма баллов по каждому пункту Е.1—Е.5. Итоговый балл находится в диапазоне [0; 14]. Приемлемой считается оценка не менее восьми баллов, при этом недопустимо отсутствие баллов хотя бы по одному пункту.

Библиография

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

[2] МКБ-10 Международная статистическая классификация болезней и проблем, связанных со здоровьем (10-й пересмотр)

[3] Федеральный закон от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации»

[4] Васильев Ю.А., Бобровская Т.М., Арзамасов К.М., Четвериков С.Ф., Владзимирский А.В., Омелянская О.В., Андрейченко А.Е., Павлов Н.А., Анищенко Л.Н. Основополагающие принципы стандартизации и систематизации информации о наборах данных для машинного обучения в медицинской диагностике // Менеджер здравоохранения. 2023. № 4 — С. 28—41

[5] Демонстрационный тестовый набор данных для разработки, дообучения и тестирования программных продуктов / ГБУЗ «НПКЦ ДиТ ДЗМ». — М., 2024

УДК 615.841:006.354 ОКС 11.040.01

Ключевые слова: искусственный интеллект, клиническая медицина, универсальность, структурированность, набор данных

Редактор *Е.В. Якубова*
Технический редактор *В.Н. Прусакова*
Корректор *М.И. Першина*
Компьютерная верстка *А.Н. Золотаревой*

Сдано в набор 29.10.2025. Подписано в печать 19.11.2025. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 2,79. Уч.-изд. л. 2,00.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

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

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

gost.e-ecolog.ru