|  |  |  |
| --- | --- | --- |
| **ФЕДЕРАЛЬНОЕ АГЕНТСТВО**  **ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ** | | |
|  | **НАЦИОНАЛЬНЫЙ СТАНДАРТ**  **РОССИЙСКОЙ**  **ФЕДЕРАЦИИ** | **ГОСТ Р**  **XXXXXX—2023 / ISO/IEC 8183:2023** |

Информационные технологии

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

**Структура жизненного цикла данных**

**(ISO/IEC 8183:2023, MOD)**

**Издание официальное**

**Москва**

**Российский институт стандартизации**

**2023**

# Предисловие

1 ПОДГОТОВЛЕН Научно-образовательным центром компетенций в области цифровой экономики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В.Ломоносова» (МГУ имени М.В.Ломоносова) и Обществом с ограниченной ответственностью «Институт развития информационного общества» (ИРИО) на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от\_\_ \_\_\_\_\_\_\_\_\_ 202\_ г. № \_\_\_\_

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 8183 «Информационные технологии – Искусственный интеллект – Структура жизненного цикла данных» (ISO/IEC 8183:2023, Information technology — Artificial intelligence — Data life cycle framework, MOD) путем изменения отдельных фраз (слов, значений).

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им национальные стандарты, сведения о которых приведены в дополнительном приложении ДА.

5 ВВЕДЕН ВПЕРВЫЕ

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

© ISO, 2023

© IEC, 2023

© Оформление. ФГБУ «РСТ», 2023

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

Содержание

[Предисловие II](#_Toc134107060)

[Введение V](#_Toc134107061)

[1 Область применения 1](#_Toc134107062)

[2 Нормативные ссылки 1](#_Toc134107063)

[3 Термины и определения 1](#_Toc134107064)

[4 Сокращения 2](#_Toc134107065)

[5 Обзор жизненного цикла данных 2](#_Toc134107066)

[6 Структура жизненного цикла данных 3](#_Toc134107067)

[6.1 Общие положения 3](#_Toc134107068)

[6.2 Этап 1. Концептуализация идеи 4](#_Toc134107069)

[6.3 Этап 2. Формирование деловых требований 5](#_Toc134107070)

[6.4 Этап 3: Планирование работы с данными 6](#_Toc134107071)

[6.5 Этап 4: Комплектование наборов данных 7](#_Toc134107072)

[6.6 Этап 5: Подготовка наборов данных 7](#_Toc134107073)

[6.7 Этап 6: Построение модели 9](#_Toc134107074)

[6.8 Этап 7: Развертывание системы 10](#_Toc134107075)

[6.9 Этап 8: Эксплуатация системы 11](#_Toc134107076)

[6.10 Этап 9: Вывод данных из эксплуатации 11](#_Toc134107077)

[6.11 Этап 10: Вывод системы из эксплуатации 12](#_Toc134107078)

[7 Этапы и процессы жизненного цикла данных 12](#_Toc134107079)

[7.1 Общие положения 12](#_Toc134107080)

[Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов национальным стандартам 16](#_Toc134107081)

[Библиография 17](#_Toc134107082)

# Введение

Системы искусственного интеллекта внедряются организациями всех типов, размеров и целей. Данные необходимы для разработки и эксплуатации систем искусственного интеллекта.

В области систем искусственного интеллекта имеется большое число жизненных циклов данных, используемых и рассматриваемых для различных целей (например, для оценки качества данных, систематических ошибок в данных, управления данными, разработки и использования систем искусственного интеллекта). Без всеобъемлющей структуры эти различные жизненные циклы данных могут быть сложны для правильной интерпретации теми, кто не имеет предварительных знаний, контекста и опыта. Существует риск того, что многочисленные жизненные циклы данных не будут применяться должным образом.

Настоящий документ содержит обзор жизненного цикла данных в разделе 5, описывает структуру жизненного цикла данных в разделе 6 и предоставляет дополнительные сведения об этапах или процессах жизненного цикла данных в разделе 7.

**НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Информационные технологии**

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

**Структура жизненного цикла данных**

Information technology. Artificial intelligence. Data life cycle framework

**Дата введения − 2023-\_\_-\_\_**

# 1 Область применения

Настоящий стандарт определяет этапы и связанные с ними действия по обработке данных на всем жизненном цикле системы искусственного интеллекта, включая комплектование, создание, разработку, развертывание, техническую поддержку и вывод из эксплуатации. Настоящий стандарт не определяет конкретные сервисы, платформы или инструменты. Данный стандарт применим ко всем организациям, независимо от их типа, размера и характера, которые используют данные при разработке и использовании систем искусственного интеллекта.

# 2 Нормативные ссылки

В настоящем стандарте использована нормативная ссылка на следующий стандарт [для датированных ссылок применяют только указанное издание ссылочного стандарта, для недатированных — последнее издание (включая все изменения)]:

ISO/IEC 22989:2022, Information technology — Artificial intelligence — Artificial intelligence concepts and terminology (Информационные технологии. Искусственный интеллект. Концепции искусственного интеллекта и терминология).

# 3 Термины и определения

В настоящем стандарте применяются термины и определения, приведенные в ISO/IEC 22989.

ИСО и МЭК поддерживают терминологические базы данных для использования в стандартизации, расположенные по следующим адресам:

* платформа ИСО для онлайн-просмотра материалов по стандартам (Online Browsing Platform, OBP) доступна по адресу <https://www.iso.org/obp/ui>;
* Электропедия МЭК (IEC Electropediа) доступна по адресу <http://www.electropedia.org/>.

# 4 Сокращения

|  |  |  |
| --- | --- | --- |
| ИИ | — | искусственный интеллект (artificial intelligence) |
| DPIA | — | оценка воздействия защиты данных (data protection impact assessment) |
| JSON | — | тестовый формат описания объектов JavaScript (JavaScript object notation) |
| МО | — | машинное обучение (machine learning) |
| OWL | — | язык веб-онтологий (web ontology language) |
| ПДн | — | персональные данные (personally identifiable information) |
| XML | — | расширяемый язык разметки (extensible markup language) |

# 5 Обзор жизненного цикла данных

Жизненный цикл данных для систем искусственного интеллекта включает в себя процессы оперирования данными от первоначальной концепции создания новой системы искусственного интеллекта до окончательного вывода этой системы из эксплуатации и разделен на несколько отдельных этапов. Каждый этап, как правило, является частью жизненного цикла данных для системы искусственного интеллекта.

Жизненный цикл данных представляет собой все этапы, через которые данные могут проходить в любой системе, оперирующей данными произвольного типа. Он предназначен для поддержки достижения целей, связанных с управлением системой, полезностью системы, качеством данных и безопасностью данных, путем обеспечения того, чтобы обработке данных уделялось должное внимание при планировании, разработке, использовании и выводе системы из эксплуатации.

Детальное предназначение и последовательность прохождения этих этапов на протяжении жизненного цикла зависят от множества факторов, включая социальные, коммерческие, организационные и технические соображения, каждый из которых может меняться или иногда сочетаться с другими этапами в течение жизненного цикла системы. Настоящий стандарт описывает следующие 10 этапов:

* этап 1 – концептуализация идеи;
* этап 2 –формирование деловых требований;
* этап 3 – планирование работы с данными;
* этап 4 – комплектование наборов данных;
* этап 5 – подготовка наборов данных;
* этап 6 – построение модели;
* этап 7 – развертывание системы;
* этап 8 – эксплуатация системы;
* этап 9 – вывод данных из эксплуатации;
* этап 10 – вывод системы из эксплуатации.

Сведения о жизненном цикле данных при их использовании см. в ISO/IEC 5212[[1]](#footnote-1).

# 6 Структура жизненного цикла данных

## 6.1 Общие положения

Структура жизненного цикла данных на рисунке 1 идентифицирует множество концептуально различных этапов оперирования данными в системе ИИ – от планирования работы с данными до вывода данных из эксплуатации. Рисунок 1 также включает концептуализацию идеи, формирование деловых требований и вывод системы из эксплуатации, которые относятся к этапам жизненного цикла системы.

Сведения о наборах данных см. в ISO/IEC 23053. С каждым этапом жизненного цикла соотносятся процессы, соответствующие определенной задаче. Процессы жизненного цикла описывают действия с данными на конкретном этапе.

Этап 9 (вывод данных из эксплуатации) и этап 10 (вывод системы из эксплуатации) относятся к выводу из эксплуатации, но этап 9 конкретно описывает, что происходит с данными (например, безопасное удаление, архивирование, перепрофилирование), а этап 10 описывает, что происходит с системой независимо от того, что происходит с обрабатываемыми данными.

Изображение выглядит как диаграмма, текст, План, Технический чертеж

Автоматически созданное описание

Рисунок 1 — Структура жизненного цикла данных

Примечание 1 — Однонаправленные стрелки на рисунке 1 изображают прямой путь по этапам жизненного цикла, а двунаправленные стрелки показывают путь с обратной связью между этапами жизненного цикла.

Примечание 2 — Верификация и валидация модели относятся к внутреннему процессу разработки, результатом которого является модель. Валидация и верификация системы относится к системе в целом в ходе ее эксплуатации.

## 6.2 Этап 1. Концептуализация идеи

Концептуализация идеи (замысел) включает процесс осознания потребности или требований к новой или модернизируемой системе ИИ. Система ИИ может использоваться как частичное или полное решение существующей или предполагаемой задачи или проблемы, с которой сталкивается организация.

Концептуализация идеи также может быть обусловлена более широкими потребностями организации (например, экономическими, техническими, стратегическими, рыночными или нормативными правовыми требованиями). В конечном счете концептуализация идеи может быть зафиксирована в виде одного или нескольких вопросов, на которые может дать ответ система ИИ. Решаемые системой ИИ вопросы целесообразно сопоставлять и согласовывать с деловыми целями и показателями.

## 6.3 Этап 2. Формирование деловых требований

На этапе формирования деловых требований могут участвовать представители одной или нескольких заинтересованных сторон, обладающих соответствующими полномочиями или влиянием, и решающих выяснить, можно ли превратить замысел в функционирующую систему и следует ли вкладывать в него дополнительные средства. Этот этап включает:

* определение амбиций проекта (например, видение, цели и стратегию);
* определение имеющихся и подлежащих приобретению активов;
* спецификацию требований к данным, ключевому элементу системы ИИ, исходя из целей и потребностей конечных пользователей;
* идентификацию факторов, влияющих на реализацию проекта, включая внутренние компетенции и знания, организационную структуру, технологии и внешние ресурсы;
* обеспечение возможности реализации проекта в соответствии с политикой и процедурами (или процессами) организации, включая такие как:
* соответствие (например, требования к защите ПДн);
* этика (например, справедливость результатов);
* культура;
* лидерство;
* процессы управления.

Этап формирования деловых требований может завершаться подготовкой заключения об осуществимости проекта.

Примечание — На этапе формирования деловых требований никакие данные не обрабатываются.

## 6.4 Этап 3: Планирование работы с данными

Этап планирования работы с данными включает решение о составе наборов данных, которые нужны для ответов на вопросы, сформулированные на этапе формирования деловых требований. Основные аспекты, которые рассматриваются на этом этапе, включают:

* существуют ли необходимые данные, доступны ли они для повторного использования, нужно ли их приобретать, собирать, преобразовывать, создавать, курировать или сочетать некоторые из перечисленных выше действий;
* требуемый объем данных;
* источник данных;
* можно ли создать синтетические, т. е. искусственные данные, для дополнения имеющихся;
* какие выходные данные будут созданы и как система будет оперировать ними;
* формат данных;
* что данные представляют;
* свойства данных, которые могут повлиять на выбор алгоритма для построения модели;
* требования к лицензированию данных;
* требования к безопасности, защите персональных данных и жизнестойкости данных;
* требования к комплектованию данных, в том числе к сбору данных;
* требования к защите данных;
* требования к срокам и способам хранения данных (например, тип, стоимость, емкость, производительность, встроенные возможности, сроки удаления).

Целью этого этапа является обеспечение того, чтобы цели, требования и потребности, установленные на этапах концептуализации идеи и формирования деловых требований, могли быть выполнены. Потенциально необходимые наборы данных могут быть идентифицированы, получены и проверены. Наборы данных могут быть внутренними или получены из общедоступных источников, государственных органов, поставщиков или сторонних организаций.

## 6.5 Этап 4: Комплектование наборов данных

Этап комплектования наборов данных включает создание или доступ к данным, определенным на этапе планирования работы с данными. Приобретенные данные могут поступать из внутренних источников, от третьих лиц или сообществ (например, открытые данные, общедоступные данные).

Получение данных от третьих лиц может основываться на согласиях, договорах или лицензиях, как описано в 6.4.

Данные могут быть в разных формах (например, статические данные, потоковые данные, данные интернета вещей в реальном времени) и в разных форматах (таких как XML, JSON, текст с разделителями, бинарный формат). Данные могут быть структурированными, полуструктурированными или неструктурированными.

В процессах комплектования наборов данных должны использоваться передовые практики управления данными (такие как безопасность, защита ПДн, качество).

## 6.6 Этап 5: Подготовка наборов данных

Этап подготовки наборов данных включает в себя обработку данных, собранных на этапе комплектования. Этап подготовки наборов данных может включать выполнение следующих операций.

* Расшифровка: преобразование зашифрованных данных в состояние, при котором их можно использовать в системе ИИ, когда это необходимо и возможно.
* Очистка: включает в себя преобразования и операции, такие как валидация релевантности, дедупликация, удаление выбросов, устранение систематических ошибок, дополнение недостающих значений, корректировка записей и исправление форматов данных.
* Определение состава функциональных возможностей: отбор подходящих функций для обработки и возможного преобразования с целью повышения производительности при машинном обучении. Существующие функции могут комбинироваться и использоваться для порождения новых функций, которые могут улучшить обучение и логические выводы.
* Нормализация и масштабирование: может возникнуть необходимость преобразования данных из сильно различающихся диапазонов в определенный диапазон (например, между 0 и 1 единицей измерения). Также может быть необходимо масштабировать выборки данных, чтобы они соответствовали стандартному распределению или другому заданному распределению.
* Организация данных: данные могут быть реорганизованы без изменения их значения или смысла. Для получения полного набора данных может понадобиться еще добавить наборы данных или объединить таблицы. Может потребоваться объединить или разделить столбцы или поля для достижения необходимой структуры набора данных.
* Разметка (маркировка): значения целевых переменных должны быть установлены с помощью подходящего ручного или автоматического процесса. Например, при обучении с учителем значения можно определять вручную, при полуконтролируемом обучении значения могут быть определены с помощью автоматизированных методов. Примером ручной разметки данных для программы распознавания изображений является использование людей при определении видов животных в множестве цифровых изображений.
* Обогащение: запуск инструментов для связывания различных источников данных и добавления дополнительного контекста к данным. Например, неструктурированные данные могут быть обогащены средствами обработки естественного языка для извлечения именованных сущностей. Названия местностей и адреса могут быть идентифицированы и геокодированы с помощью географического справочника, чтобы впоследствии можно было проводить анализ на основе местоположения;
* Обезличивание: может возникнуть необходимость для удаления персональных данных из наборов данных для защиты неприкосновенности частной жизни субъектов данных.
* Повторная выборка: например, может быть полезно использовать частичную выборку из больших наборов данных, чтобы улучшить согласованность статистической значимости различных классов данных или сократить время, необходимое для построения и тестирования модели, при этом получая полезные результаты. Точно так же можно пополнить выборку (т. е. расширить с заменой данных), чтобы повысить согласованность статистической значимости различных классов данных.
* Кодирование: может потребоваться кодирование данных, используемых для построения модели. Например, может потребоваться кодирование текстовых значений для категориальных переменных (такое как преобразование текстовых признаков в числовые или оцифровка аналоговых сигналов).
* Верификация целостности: применение процесса, специфичного для конкретного типа данных с целью проверки общей целостности набора данных. Скорее всего, это применимо к структурированным и частично структурированным данным, для которых уже может существовать структурная модель (например, схема базы данных, формальная онтология).
* Происхождение данных: обновление записи происхождения каждого набора данных для записи изменений и предпринятых операций.
* Анонимизация данных или псевдонимизация.

Примечание — Дополнительную информацию о подготовке наборов данных для машинного обучения см. ISO/IEC 23053.

## 6.7 Этап 6: Построение модели

Построение модели включает принятие решения об организации, хранении и доступе к данным таким образом, чтобы их можно было обработать для построения модели, выполняющей отдельные функции для достижения целей. Процесс построения модели может быть отдельным действием с выводом, результатом которого является фиксированная модель, или непрерывным действием, при котором модель подвергается постоянному пересмотру (непрерывное обучение). В любом случае модель может возникнуть как:

1. Результат обучения алгоритма машинного обучения с использованием обучающих данных. Построение модели может происходить централизованно или через сеть ресурсов (например, федеративное обучение, раздельное обучение). Примеры обученных моделей включают деревья решений, индуктивное логическое программирование и различные типы нейронных сетей.
2. Результат объединения инженерно-технических знаний (например, декларативных или процедурных) с процессом логического вывода. Примеры форм инженерных знаний человека включают предложения Хорна (например, используемые в языке программирования Prolog), разновидности логики описания (например, используемые в языке веб-онтологий OWL и OWL2) и программирование набора ответов.

Примечание 1 — Индуктивное логическое программирование — это форма машинного обучения символьных структур, в которой логическая программа автоматически модифицируется для удовлетворения заданным целевым условиям.

Примечание 2 — Программирование набора ответов — это форма логического программирования, в которой используется алгоритм решения набора ответов для построения символической модели, в которой все переменные заменены литералами.

Данные используются для обучения и калибровки модели наряду с человеческим опытом, а также для верификации того, что результаты функционирования и производительность системы соответствуют ожиданиям заинтересованных сторон. Оценка воздействия обработки данных (Data Processing Impact Assessment, DPIA) может быть выполнена для решения таких вопросов, как соблюдение требований к защите персональных данных в полученных результатах. Модель также может быть оценена на наличие других потенциальных проблем (таких как предвзятость, справедливость, другие этические проблемы) и предпринятых корректирующих действий.

На этапе построения следует оценить модель и убедиться, что она соответствует требованиям, установленным на предыдущих этапах (таких как формирование деловых требований, планирование работы с данными, комплектование наборов данных). Аналогичным образом результаты и эффективность модели следует оценивать в сравнении с ожиданиями соответствующих заинтересованных сторон, включая их способность использовать модель на практике. В некоторых случаях для систем, связанных с общественной безопасностью, может потребоваться независимая оценка безопасности (Independent Safety Assessment, ISA).

## 6.8 Этап 7: Развертывание системы

Развертывание системы включает в себя «запуск» системы искусственного интеллекта в целевой среде. Данный этап не обязательно представляет собой простое включение системы; вместо этого он может включать ряд процессов, гарантирующих, что система работает должным образом. В рамках этапа необходимо проверить потоки данных, чтобы убедиться, что они работают так, как предполагалось, особенно если целевая среда включает новые системы или соединения.

## 6.9 Этап 8: Эксплуатация системы

Эксплуатация системы включает использование модели, генерирующей выходные данные из входных (или произведенных) данных. Выходные данные также могут быть обработаны несколькими способами, такими как:

* прием или извлечение данных;
* псевдонимизация или анонимизация данных;
* манипулирование данными или комбинирование данных;
* анализ данных;
* визуализация данных;
* передача данных;
* хранение данных.

Этот этап может включать авторизацию доступа к данным, аутентификацию и предполагаемое использование.

Произведенные в ходе эксплуатации данные должны постоянно контролироваться, чтобы гарантировать, что качество данных поддерживается, а система не используется для целей, которые изначально не предполагались.

Непрерывная верификация и проверка могут быть полезны для снижения рисков при работе системы, но не всегда осуществимы или уместны. Когда это возможно и уместно, следует проводить постоянную верификацию и проверку системы, чтобы гарантировать выполнение деловых требований и соответствующих ожиданий заинтересованных сторон. Система может постоянно улучшаться по мере необходимости. Новые обучающие данные можно использовать для смягчения снижения производительности. Управлением новыми обучающими данными следует заниматься на этапе 4 (комплектование наборов данных) и этапе 5 (подготовка наборов данных).

## 6.10 Этап 9: Вывод данных из эксплуатации

Вывод данных из эксплуатации предусматривает решение судьбы данных, которые больше не используются системой (например, их безопасное удаление, архивирование, перепрофилирование). Категории данных должны быть определены, а некоторые категории данных должны быть сохранены для целей аудита (например, данные журналирования для подтверждения соответствия).

Модель тоже может быть выведена из эксплуатации, если она содержит остаточные элементы обучающих данных или из-за других требований (например, безопасности, приватности и конфиденциальности). Требования лицензирования данных могут предполагать удаление данных определенного уровня. Кроме того, может потребоваться удаление ПДн, например, в соответствии с требованиями законодательства или условиями договора.

## 6.11 Этап 10: Вывод системы из эксплуатации

Вывод системы из эксплуатации включает в себя прекращение обработки данных и утилизацию компонентов системы, на которые не распространяется вывод данных из эксплуатации, таких как компоненты целевой среды. Данные, не относящиеся к модели, такие как системные журналы, можно сохранить для дальнейшего изучения или безопасно удалить.

# 7 Этапы и процессы жизненного цикла данных

## 7.1 Общие положения

Изображение выглядит как текст, чек, число, диаграмма

Автоматически созданное описаниеНа рис. 2 показаны процессы, важные для жизненного цикла данных.

Рисунок 2 — Наглядная демонстрация того, как процессы соотносятся с каждым этапом

На рис. 2 показаны процессы и сквозные аспекты, важные для жизненного цикла данных. К ним относятся следующие процессы.

* Валидация и верификация: два вопроса «Построили ли мы правильную систему?» и «Построили ли мы систему правильно?» постоянно проверяются путем тестирования выходных результатов системы на соответствие деловым требованиям.
* Бизнес-анализ: взаимосвязь между требованиями организации и способностью системы ИИ и особенно используемых в ней данных, соответствовать этим требованиям анализируется на различных этапах жизненного цикла данных.
* Обработка данных: под обработкой понимаются любые операции с данными, включая сбор, сохранение, передачу, удаление или иные действия. Это повышает значимость персональных данных и особенно специальных категорий персональных данных. Сведения о происхождении данных, используемых в системах ИИ, должны актуализироваться в соответствии с требованиями организации.
* Безопасность данных: данные в системе ИИ должны быть конфиденциальными (т. е. защищенными от несанкционированного доступа), доступными (т. е. доступными для санкционированного доступа) и целостными (т. е. защищенными от несанкционированного изменения).
* Неприкосновенность частной жизни: ПДн должны быть защищены, а их целостность и доступность должны поддерживаться на основе принципов конфиденциальности, целостности, доступности. Процессы должны включать защиту ПДн, а также предотвращать случайное или преднамеренное создание новых ПДн.
* Безопасность: убедитесь, что эффективность и результативность оперирования данными системой снижают потенциальный риск или вред, которые могут быть порождены системой ИИ.

***Пример 1. Разметка данных: система ИИ, которая обучается на наборе тестовых данных для корректной разметки визуального контента для ограничения просмотра по возрасту. Если система ИИ не была обучена избегать некорректной идентификации элементов визуального контента по возрасту, то в процессе эксплуатации она подвергает будущую аудиторию риску.***

***Пример 2. Мониторинг данных: непрерывный мониторинг данных, которые вводятся для анализа, необходим для выявления изменений в данных, которые не могут быть обработаны в процессе анализа. Если начальные условия неверны, то возможными исходами работы системы станут незавершенность или бессмысленные результаты. Анализ может быть ошибочным, потому что система оперирует данными, которые не удовлетворяют свойствам, которые она должна была учитывать.***

***Пример 3. Нормативные правовые требования и обеспечение их исполнения: если данные могут использоваться в приложениях ИИ, на которые распространяются конкретные требования законодательства или требования общественной безопасности (например, в железнодорожной, авиационной, атомной, медицинской или нефтегазовой отраслях), то к данным могут возникнуть дополнительные требования.***

Примечание — Исполнение требований к наборам данных согласовывается с регулирующим органом или оценивается им. Это может быть сделано как в рамках оценки функциональной безопасности, так и в более широком контексте. Требования могут применяться как для статических данных, так и для текущих или динамических данных.

* Разработка системы: создание системы, которую можно использовать для выполнения деловых требований организации, включая их непрерывную валидацию и верификацию.
* Управление рисками: идентификация и управление любыми рисками, связанными с системой, на основе склонности организации к риску. Сюда включены риски, связанные с защитой персональных данных и безопасностью системы и данных в ней.
* Стратегическое управление: роль руководящего органа и менеджеров в эксплуатации и выводе из эксплуатации системы искусственного интеллекта и данных, используемых в ней. Это включает рассмотрение этических проблем, соблюдение требований законодательства, стандартов и передовой практики, подотчетности, управления рисками и фидуциарных обязательств (см. ISO/IEC 38507).
* Социальная и экологическая устойчивость: учитываются социальные и экологические последствия разработки и использования системы ИИ, чтобы данные, обрабатываемые в системе, можно было обрабатывать максимально устойчивым образом. Это относится как к физическому функционированию системы (например, путем обеспечения возможности работы в периоды низкого энергопотребления или предотвращения ненужного использования системы), так и к результатам функционирования системы (таким как соблюдение этических норм).

# Приложение ДА (справочное) Сведения о соответствии ссылочных международных стандартов национальным стандартам

Таблица ДА.1

|  |  |  |
| --- | --- | --- |
| Обозначение ссылочного международного стандарта | Степень соответствия | Обозначение и наименование соответствующего национального стандарта |
| ISO/IEC 22989:2022, Information technology — Artificial intelligence — Artificial intelligence concepts and terminology | NEQ\* | ПНСТ 553–2021\*\* Информационные технологии. Искусственный интеллект. Концепции искусственного интеллекта и терминология (ISO/IEC DIS 22989, NEQ) |
| \* В настоящей таблице использовано следующее условное обозначение степени соответствия стандарта:  NEQ — неэквивалентный стандарт.  \*\* В России принят предварительный национальный стандарт, основанный на ранней редакции международного стандарта | | |

# Библиография

|  |  |  |
| --- | --- | --- |
| [1] |  | ISO/IEC 5212[[2]](#footnote-2), Information technology — Data usage — Guidance for data usage |
| [2] |  | ISO/IEC 23053, Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) |
| [3] |  | ISO/IEC 38507, Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations |

УДК 004.01:006.354 ОКС 35.020

Ключевые слова: информационные технологии, искусственный интеллект, рамочная структура, жизненный цикл, жизненный цикл данных, большие данные, набор данных, этап жизненного цикла, процесс жизненного цикла, модель данных, комплектование наборов данных, подготовка наборов данных, вывод данных из эксплуатации, жизненный цикл системы искусственного интеллекта

|  |  |  |
| --- | --- | --- |
| Руководитель разработки  Председатель совета директоров  ООО «Институт развития  информационного общества» |  | Ю. Е. Хохлов |

1. В разработке. На момент публикации: ISO/IEC DIS 5212:2023 [↑](#footnote-ref-1)
2. В разработке. На момент публикации: ISO/IEC DIS 5212:2023 [↑](#footnote-ref-2)