ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 71539— 2024 (ИСО/МЭК 5338:2023)

Искусственный интеллект

ПРОЦЕССЫ ЖИЗНЕННОГО ЦИКЛА СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

(ISO/IEC 5338:2023, Information technology — Artificial intelligence — AI system life cycle processes, MOD)

Издание официальное

Москва
Российский институт стандартизации
2024

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 28 октября 2024 г. № 1539-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 5338:2023 «Информационные технологии. Искусственный интеллект. Процессы жизненного цикла системы искусственного интеллекта» (ISO/IEC 5338:2023 «Information technology — Artificial intelligence — AI system life cycle processes», MOD) путем изменения отдельных фраз (слов, значений показателей, ссылок), которые выделены в тексте курсивом.

Внесение указанных технических отклонений направлено на учет особенностей национальной стандартизации технологий работы с большими данными и искусственного интеллекта.

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2021 (пункт 3.5).

Сведения о соответствии ссылочных национальных стандартов международным стандартам, используемым в качестве ссылочных в примененном международном стандарте, 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)

Содержание

Введение

Системы искусственного интеллекта добились значительных успехов в таких областях, как компьютерное зрение и распознавание изображений, обработка естественного языка, выявление мошенничества, управление беспилотными транспортными средствами, прогнозная техническая поддержка и планирование. Эффективным подходом к разработке и использованию систем искусственного интеллекта является расширение состава процессов жизненного цикла традиционной информационной системы посредством включения в него характерных для искусственного интеллекта особенностей жизненного цикла.

Примером такой специфической особенности жизненного цикла системы искусственного интеллекта является ситуация, когда в системе применяется машинное обучение с использованием обучающих данных и возникает необходимость повторно обучить модель машинного обучения на основе новых обучающих данных, которые лучше отражают особенности текущих эксплуатационных данных.

Международный стандарт *[1]* описывает процессы жизненного цикла программного обеспечения, а *[2]* — процессы жизненного цикла системы. Хотя эти процессы жизненного цикла в целом применимы к системам искусственного интеллекта, чтобы учесть их особенности, требуется ввести нескольких новых и модифицировать ряд существующих процессов.

Настоящий стандарт расширяет существующие международные стандарты типичного жизненного цикла таким образом, чтобы сделать их применимыми к системам искусственного интеллекта и жизненный цикл таких систем мог выиграть от применения устоявшихся моделей и имеющегося опыта. Некоторые системы искусственного интеллекта используются в областях, связанных с безопасностью, таких как здравоохранение или управление дорожным движением. Такие критически важные с точки зрения безопасности системы искусственного интеллекта требуют особого внимания и обсуждения, как это описано в *[3]*.

Как объясняется в *ГОСТ Р 71476*, интеграция жизненного цикла системы искусственного интеллекта в существующие процессы обеспечивает повышение эффективности, лучшее внедрение и взаимопонимание между заинтересованными сторонами. Такой интегрированный подход к жизненному циклу учитывает тот факт, что системы искусственного интеллекта обычно представляют собой комбинацию элементов, использующих технологии искусственного интеллекта, и традиционных элементов, таких как исходный код и базы данных.

Настоящий стандарт содержит дополнительные сведения о процессах жизненного цикла системы искусственного интеллекта, которые обсуждаются в стандарте *[4]*.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Искусственный интеллект

ПРОЦЕССЫ ЖИЗНЕННОГО ЦИКЛА СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Artificial intelligence. AI system life cycle processes

Дата введения — 2025—01—01

1 Область применения

Настоящий стандарт определяет набор процессов и связанных с ними понятий для описания жизненного цикла систем искусственного интеллекта на основе машинного обучения и эвристических систем. Он основан на международных стандартах *[1]* и *[2]* с модификациями и добавлением специфических для искусственного интеллекта процессов по *ГОСТ Р 71476* и *[5]*.

В настоящем стандарте описаны процессы, поддерживающие определение, контроль, управление, функционирование и совершенствование системы искусственного интеллекта на стадиях ее жизненного цикла. Эти процессы также могут быть использованы в рамках организации или проекта при разработке или приобретении систем искусственного интеллекта. В тех случаях, когда элементом системы искусственного интеллекта является традиционное программное обеспечение или традиционная информационная система, при реализации такого элемента можно использовать процессы жизненного цикла программного обеспечения в соответствии с *[1]* и процессы жизненного цикла системы в соответствии с *[2]*.

2 Нормативные ссылки

В настоящем стандарте использована нормативная ссылка на следующий стандарт:

ГОСТ Р 71476—2024 (ИСО/МЭК 22989:2022) Искусственный интеллект. Концепции и терминология искусственного интеллекта

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.

Издание официальное

3 Термины и определения

В настоящем стандарте применены термины по *ГОСТ Р 71476*, *[1]*, *[2]*, *[5]*, а также следующий термин с соответствующим определением:

3.1

приобретение знаний (knowledge acquisition): Процесс определения местонахождения, сбора и уточнения знаний, а также преобразование их к виду, который может в дальнейшем обрабатываться системой, основанной на знаниях.

Примечание — Приобретение знаний обычно подразумевает участие инженера по знаниям, однако оно также является важным элементом машинного обучения.

[ГОСТ 33707—2016, пункт 4.1065]

4 Сокращения

В настоящем стандарте применены следующие сокращения:

ИИ — искусственный интеллект;

МО — машинное обучение.

5 Основные понятия

5.1 Общие положения

Настоящий стандарт определяет элементы процессов, характерные для жизненного цикла системы ИИ.

Жизненный цикл системы ИИ состоит из процессов трех типов:

- типовые процессы: процессы, идентичные тем, что определены в *[1]*, *[2]*;
- модифицированные процессы: процессы, отдельные элементы которых были изменены, добавлены или удалены по сравнению с их определениями в *[1]*, *[2]*;
- процессы, специфические для ИИ: процессы, являющиеся специфическими для систем ИИ и не основанные непосредственно на каких-либо процессах, определенных в *[1]*, *[2]*.

Примечание — Раздел о каждом «модифицированном процессе» включает подраздел, посвященный характерным для ИИ особенностям, в котором содержатся рекомендации по адаптации процесса к системам ИИ.

Процессы жизненного цикла системы ИИ в разделе 6 представлены как типовые, модифицированные либо специфические для ИИ. На рисунке 1 показаны процессы жизненного цикла системы ИИ, сгруппированные по типам и сопоставленные с *[2, рисунок 4]*.

Перечисленные ниже аспекты систем ИИ являются ключевыми факторами, отличающими процессы их жизненного цикла от процессов жизненного цикла традиционных систем:

- измеримая потенциальная деградация: поскольку модели ИИ нацелены на моделирование желаемого поведения, которое может изменяться со временем, то могут понадобиться измерения и мониторинг любых отклонений в эксплуатационных данных (дрейф данных — data drift) и отклонений, влияющих на целевой результат (дрейф концепции — concept drift). Изменение желаемого поведения возможно не только у систем ИИ, однако для моделей ИИ оно является однозначно измеримым путем валидации входных и выходных данных;
- потенциальная автономность: способность системы ИИ автоматически и быстро принимать сложные решения создает потенциал для замены ею действий и процессов, которые в противном случае выполнялись бы людьми. В этой связи может потребоваться дополнительное внимание к системам ИИ с целью обеспечения справедливости, безопасности, защищенности, неприкосновенности частной жизни и защиты персональных данных, надежности, прозрачности и объяснимости, подотчетности, доступности, целостности и сопровождаемости. Чем выше вероятность того, что система ИИ способна причинить вред, тем важнее становится это дополнительное внимание. Обзор этических и социальных проблем при разработке и развертывании систем ИИ см. в техническом отчете *[6]*. Для получения дополнительной информации об управлении рисками систем ИИ см. *[7]*;

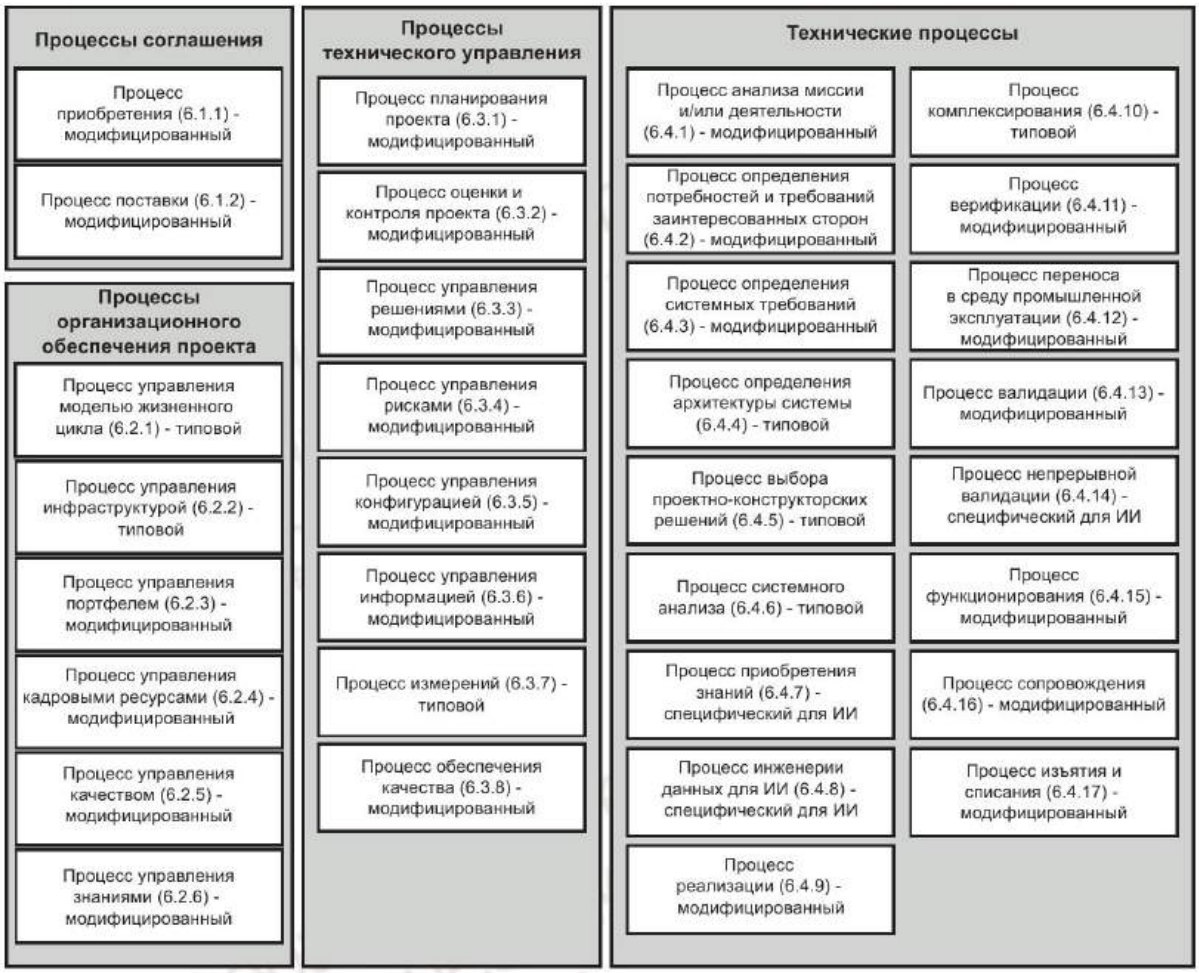

Рисунок 1 — Процессы жизненного цикла системы ИИ, сопоставленные с *[2, рисунок 4]*

- итеративное специфицирование требований и поведения: системы ИИ могут основываться на итеративных и гибких спецификациях требований, на спецификациях знаний, на моделировании поведения и на проектировании с учетом удобства использования. Разработка системы ИИ может проходить через циклы спецификации требований, создания демонстрационного прототипа и уточнения требований. В этом аспекте системы ИИ отличаются от традиционных программных приложений, основанных на фиксированных, четко сформулированных требованиях. Кроме того, в ходе использования систем ИИ требования также могут эволюционировать по мере возникновения непредвиденных ситуаций и по мере выявления уточненных требований, спецификаций и пробелов;

- вероятностный характер: решения, принимаемые системами ИИ, основанными на МО, по своей природе носят вероятностный характер. В этой связи заинтересованным сторонам важно понимать, что принятые системами ИИ решения не всегда будут правильными. Формальное тестирование правильности моделей имеет свои внутренне присущие ограничения и неопределенности, когда речь идет о гарантиях;

- зависимость от данных: основанные на МО системы ИИ при проведении обучения, тестирования и валидации моделей полагаются на достаточные, репрезентативные данные. Поведение моделей МО не программируется, а «выучивается» из данных. По этой причине важно уделять особое внимание данным (например, их качеству), которые требуются системе ИИ для обучения, тестирования, верификации и валидации;

- интенсивное использование знаний: для эвристических моделей сравнительно большое значение имеет приобретение знаний, поскольку знания явно кодируются в модели и определяют ее правильность;

- новизна: организациям, которые проектируют, разрабатывают или используют системы ИИ, могут потребоваться новые знания и навыки. Другие заинтересованные стороны, такие как пользователи систем ИИ, могут быть не знакомы с ИИ, что способно вызвать проблемы с доверием и внедрением. Новизна ИИ может стать причиной чрезмерной уверенности и энтузиазма при отсутствии полного учета рисков системы ИИ. Представления о том, что системы ИИ смогут в конечном счете заменить людей или продемонстрировать свою «разумность», также могут повлиять на то, как заинтересованные стороны смотрят на системы ИИ;

- непредсказуемость: в случае использования эвристических моделей или МО поведение модели является заранее непредопределенным, «возникающим» (эмерджентным) в том смысле, что оно не программируется явно, а является косвенным результатом инженерии знаний или же выводится из обучающих данных. Заинтересованные стороны могут обнаружить, что системы ИИ менее предсказуемы, объяснимы, прозрачны, робастны и понятны, чем явно запрограммированные системы. Это может снизить доверие к системам ИИ.

Примечание — Высокоуровневый обзор этических и социальных проблем ИИ представлен в техническом отчете *[6]*. Дополнительная информация о решении этических проблем при проектировании системы приведена в *[8]*.

5.2 Понятия, относящиеся к системе ИИ

Модель может быть либо моделью МО, обученной выполнять вычисления на основе данных, либо эвристической моделью, спроектированной на основе человеческих знаний (инженерия знаний). В эвристической модели выполнение вычислений организуется либо явным образом (процедурные), либо неявно, посредством указания правил или вероятностей (декларативные), либо используются оба подхода вместе.

В случае МО модель в первую очередь создается на основе данных, а в случае эвристической модели — на основе знаний. Как бы то ни было, в любом случае требуются как данные, так и знания. Данные необходимы для тестирования эвристических моделей и выполнения анализа с целью получения знаний. Знания же необходимы для понимания контекста, в котором используется модель МО, а также для помощи в отборе и подготовке данных для обучения и тестирования.

Для традиционных систем также часто важны как знания, так и данные. Знания могут потребоваться для реализации бизнес-логики. Данные обычно играют важную роль в любой системе обработки данных и могут потребоваться для функционального тестирования.

Различие между системой ИИ и приложением ИИ объясняется в *[9]*. Отличительные признаки приложений ИИ также определены в *[9]*.

5.3 Модель жизненного цикла системы ИИ

Модель жизненного цикла системы ИИ описывает эволюцию системы ИИ от возникновения замысла до вывода ее из эксплуатации. Настоящий стандарт не предписывает какого-либо конкретного жизненного цикла. Вместо этого в нем основное внимание обращается на характерные для ИИ процессы, которые могут происходить в течение жизненного цикла системы. Характерные для ИИ процессы могут происходить на одной или нескольких стадиях жизненного цикла, а отдельные стадии жизненного цикла могут повторяться в течение существования системы. Например, возможно, что на стадии повторной оценки разработка и развертывание будут неоднократно повторяться с целью разработки и внедрения исправлений ошибок и обновлений системы.

Модель жизненного цикла системы помогает заинтересованным сторонам создавать системы ИИ более эффективно и продуктивно. Для разработки модели жизненного цикла полезны международные стандарты, в том числе стандарты *[2]* для систем в целом, *[1]* — для программного обеспечения и *[10]* — для документации на систему. Эти международные стандарты описывают процессы жизненного цикла для традиционных систем. Рисунок 2 основан на рисунке 3 по *ГОСТ Р 71476—2024*. Он показывает пример стадий и высокоуровневых процессов, которые могут применяться при разработке и на протяжении жизненного цикла систем ИИ. Подробности см. в *ГОСТ Р 71476—2024* (пункт 6.1*)*.

Быть владельцами и управлять жизненным циклом системы ИИ или любого подмножества его стадий (таких, например, как сбор и предоставление данных, модели МО или кода для других компонентов, используемых для разработки или развертывания системы ИИ, могут разные организации и/или субъекты. Помимо этого, организация может зависеть от других организаций при создании инфраструктуры или при обеспечении необходимых возможностей жизненного цикла системы ИИ (примером может служить

создание инфраструктуры в локальной, облачной или гибридной средах). В настоящем стандарте принимаются во внимание последствия, особенности и связанные с ними риски цепочки поставок системы ИИ с тем, чтобы предложить новые, а также адаптировать и приспособить существующие процессы для создания системы ИИ, пересекающей границы организации.

Кроме того, для некоторых областей существуют специальные международные стандарты жизненного цикла, например для медицинских устройств, в отношении которых действует *[11]*. Организации должны рассматривать описанные в настоящем стандарте специфические особенности ИИ совместно с *[11]* при внедрении характерных для конкретной предметной области стандартов.

Рисунок 2 — Пример стадий и высокоуровневых процессов в модели жизненного цикла системы ИИ

Стадии на рисунке 3 основаны на стадиях, описанных в *ГОСТ Р 71476*; они показаны вместе с группами описанных в настоящем стандарте технических процессов. Стадия «непрерывная валидация» не имеет пометки «в случае использования непрерывного обучения» в отличие от примера модели жизненного цикла в *ГОСТ Р 71476—2024*, рисунок 4. Стадия непрерывной валидации также применима в ситуациях без непрерывного обучения — например, для выявления дрейфа данных, дрейфа концепции или для обнаружения технических сбоев.

Концепция стадий предназначена для группировки имеющих определенный хронологический порядок видов деятельности для того, чтобы показать их зависимости, однако полное разделение видов деятельности во времени или в организации не предлагается. Например, при использовании гибкой методологии разработки программного обеспечения разработка и эксплуатация являются отдельными стадиями, которые выполняются одновременно. Тем не менее определенная функциональная возможность должна быть сначала реализована, прежде чем ее можно будет проверить, а затем развернуть.

Кроме того, последовательность прохождения стадий может идти против направления стрелок — например, когда после стадии верификации и валидации принимается решение о повторном выполнении определенных действий в рамках стадии проектирования и разработки.

Примечание — У стадий жизненного цикла, показанных на рисунке 3, могут иметься критерии входа и выхода, основанные на специфических требованиях рассматриваемой системы (см. *[12]*).

Модель ИИ может быть либо моделью МО, либо эвристической моделью.

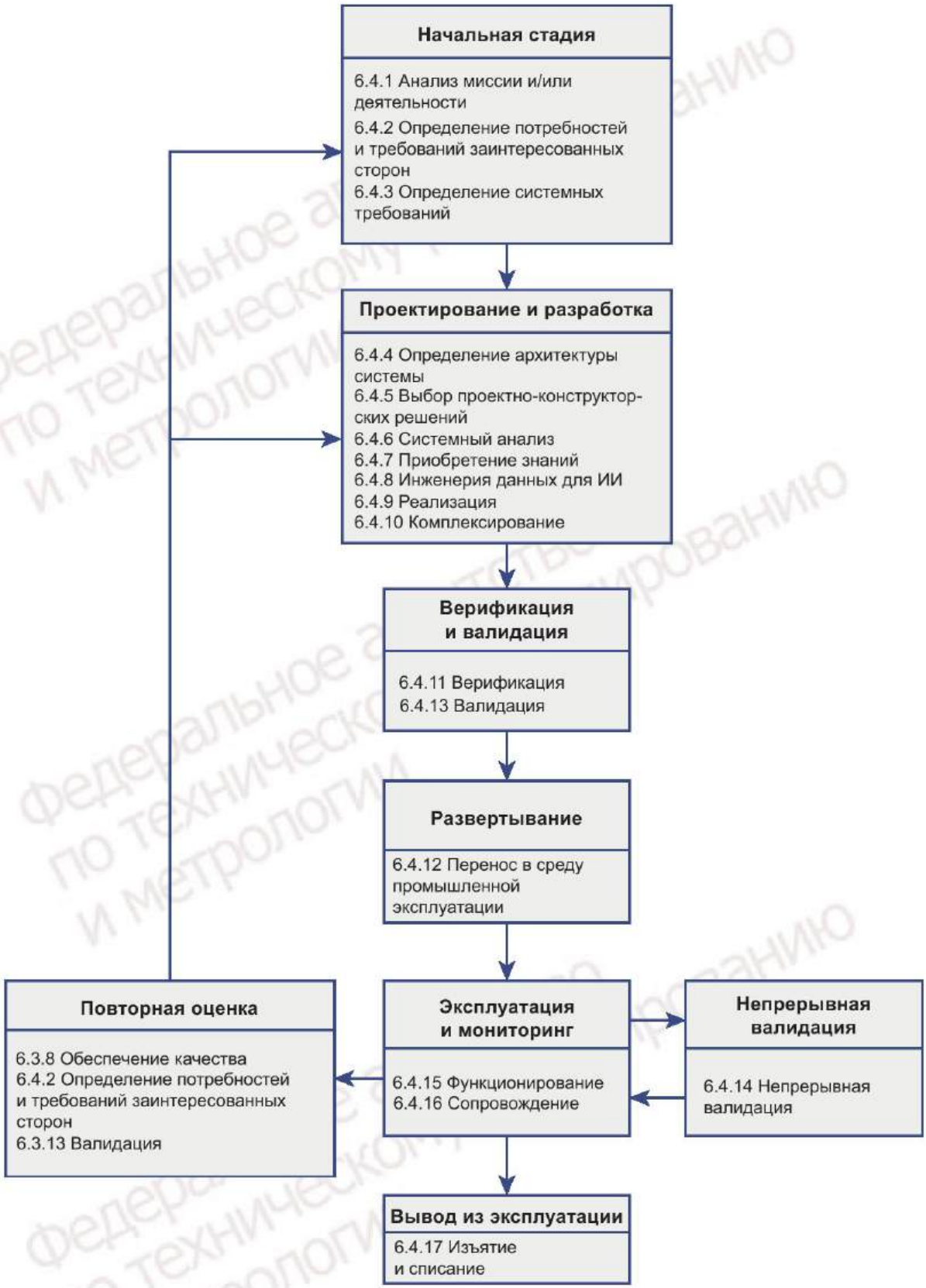

Рисунок 3 — Стадии жизненного цикла системы ИИ (с техническими процессами)

Ключевые технические процессы разработки моделей МО интегрированы в процессы жизненного цикла следующим образом:

- процесс определения системных требований: устанавливаются требования к модели;
- процесс инженерии данных для ИИ: осуществляется сбор и обновление данных;
- процесс инженерии данных для ИИ: осуществляется подготовка данных;
- процесс реализации и процесс сопровождения: (повторно) обучается и настраивается модель;
- процесс верификации: модель тестируется перед развертыванием;
- процесс переноса в среду промышленной эксплуатации: выполняется развертывание модели;
- процесс непрерывной валидации: модель тестируется после развертывания.

Для эвристических моделей ключевые шаги интегрированы следующим образом:

- процесс определения системных требований: устанавливаются требования к модели;
- процесс приобретения знаний: приобретаются знания;
- процесс реализации и процесс сопровождения (технической поддержки): осуществляется создание и обновление модели;
- процесс верификации: модель тестируется перед развертыванием;
- процесс переноса в среду промышленной эксплуатации: выполняется развертывание модели.

Примечание — Окончательное решение о том, разрабатывать ли систему ИИ или же традиционную систему, является результатом начальной стадии, на которой учитываются требования, риски, деловые потребности и потребности заинтересованных сторон.

Приложение А содержит анализ результатов применения процессов жизненного цикла традиционных систем в вариантах использования систем ИИ, описанных в *[13]*.

5.4 Понятия, связанные с процессами

5.4.1 Критерии для процессов

Процессы жизненного цикла в настоящем стандарте основаны на тех же принципах, что и в *[1]* и *[2]*. Процессы в настоящем стандарте демонстрируют сильную взаимосвязь между их результатами, действиями и задачами. Кроме того, их описание сводит к минимуму зависимости между процессами и обеспечивает возможность выполнения процесса как одной, так и несколькими организациями. Это критически важно в связи с тем, что системы ИИ могут разрабатываться несколькими организациями и/или требовать наличия способности поддерживать их от цепочек поставок нескольких организаций.

5.4.2 Описание процессов

Описание цели процесса сохраняется в неизменном виде, если соответствующий процесс взят из *[1]* или *[2]*. В подпункте «результаты процесса» описаны результаты успешной реализации процесса. Подпункт «действия и задачи» описывает реализацию процесса в соответствии с применимыми политиками и процедурами организации. Характерные для ИИ особенности процессов из *[1]* или *[2]* описаны в подпункте с заголовком «Особенности, характерные для ИИ».

5.4.3 Соответствие настоящему стандарту

Соответствие настоящему стандарту определяется как реализация всех указанных в нем процессов, действий и задач. Если какой-либо процесс, действие или задача не актуальны для системы ИИ, то отсутствие этого процесса, действия или задачи должно быть обосновано и задокументировано. Также должны применяться требования 4.2 и 4.3 *[1]* и 4.2 и 4.3 *[2]*.

6 Процессы жизненного цикла систем ИИ

6.1 Процессы соглашения

6.1.1 Процесс приобретения

6.1.1.1 Цель

Цель процесса приобретения заключается в получении продукта или услуги в соответствии с потребностями приобретающей стороны.

Примечание — Понятие «приобретающая сторона» ссылается на роль заинтересованной стороны «заказчик ИИ», а понятие «поставщик» — на роли «производитель ИИ» и «поставщик ИИ», как они описаны в *ГОСТ Р 71476*.

6.1.1.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.1.1 *[1]* и 6.1.1 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.1.1.3 Особенности, специфические для ИИ

Процесс приобретения, описанный в *[1]* и *[2]*, следует расширить за рамки приобретения продуктов и услуг, чтобы он также охватывал возможное приобретение данных для процесса инженерии данных для ИИ (см. 6.4.8). Этот новый подвид деятельности по приобретению может привести к новым проблемам приобретения, таким как затраты, зависимости, обеспечение непрерывности, обеспечение доступности, а также проблемы с правами на данные, правилами и правовыми требованиями в отношении использования приобретенных данных. Например, важными вопросами являются заключение договоров и приемка обучающих данных, потому что заключение договоров и приемка наборов данных очень трудно формализуются. Кроме того, за действиями по приобретению могут последовать итерации действий по разработке и/или переобучению, выполняемые параллельно с функционированием системы с тем, чтобы принятый набор данных продолжал соответствовать оперативным и деловым требованиям.

6.1.2 Процесс поставки

Применимы положения 6.1.2 *[1]* и 6.1.2 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

6.1.2.1 Цель

Цель процесса поставки заключается в получении приобретающей стороной продукта или услуги, которые удовлетворяют согласованным требованиям в договоре (соглашении).

Примечание — Понятие «приобретающая сторона» ссылается на роль заинтересованной стороны «заказчик ИИ», а понятие «поставщик» — на роли «производитель ИИ» и «поставщик ИИ», как они описаны в *ГОСТ Р 71476*.

6.1.2.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.1.2 *[1]* и 6.1.2 [2], касающиеся выполняемых действий, задач и результатов процесса.

6.1.2.3 Особенности, характерные для ИИ

Для процесса поставки дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.1.2.2, поставщику следует принять во внимание упомянутые ниже особенности, характерные для ИИ, чтобы предложить провести переговоры и согласовывать с приобретателем системы ИИ следующее:

- проведение апробации (подтверждения работоспособности) концепции для инициирования разработки системы ИИ перед развертыванием;
- предоставление, сбор или приобретение достаточных для МО наборов данных;
- мониторинг системы ИИ во время ее эксплуатации на тот случай, если качество работы системы начнет меняться в зависимости от эксплуатационных (производственных) данных МО, и/или принятие мер по прекращению подобной неблагоприятной ситуации;
- анализ и улучшение системы ИИ с целью устранения любых отклонений от требуемых эксплуатационных характеристик.

6.2 Процессы организационного обеспечения проекта

6.2.1 Процесс управления моделью жизненного цикла

Применимы положения 6.2.1 *[1]* и 6.2.1 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

Примечание — Типичная модель жизненного цикла систем ИИ описана в 5.3.

6.2.2 Процесс управления инфраструктурой

Применимы положения 6.2.2 *[1]* и 6.2.2 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

6.2.3 Процесс управления портфелем проектов

6.2.3.1 Цель

Цель процесса управления портфелем проектов заключается в инициации и поддержке необходимых, достаточных и релевантных проектов, направленных на достижение стратегических целей организации. Данный процесс обеспечивает инвестирование организацией адекватных финансовых средств и ресурсов, а также получение согласия на предоставление полномочий, необходимых для осуществле-

ния отобранных проектов. В рамках данного процесса регулярно проводятся оценки, подтверждающие, что проекты оправдывают (или могут быть перенацелены таким образом, чтобы оправдывать) продолжение инвестирования.

6.2.3.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.2.3 *[1]* и 6.2.3 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.2.3.3 Особенности, специфические для ИИ

Для процесса управления портфелем проектов дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.2.3.2, организациям следует принять во внимание следующие специфические для ИИ особенности:

- в рамках определения и утверждения проектов ИИ потенциально может открыть новые возможности и обеспечить новые деловые возможности для инноваций через новый проект;
- при определении потребностей в ресурсах и выделении ресурсов для новых проектов следует учитывать, что ИИ требует специальных знаний и компетенций (см. 6.2.4);
- может быть полезно — особенно в тех случаях, когда ИИ является новым инструментом для организации, — выявить аспекты, характерные для целого ряда проектов, тогда появится возможность выработать типовой подход на основе повторного использования общих элементов или платформ систем ИИ и обмена знаниями между командами проектов;
- при оценке проектов в составе портфеля следует учитывать специфические для ИИ риски (см. 6.3.4) и особенности, касающиеся планирования проектов. Например, для экспериментирования с целью обучения приемлемых моделей МО может потребоваться длительное время.

6.2.4 Процесс управления кадровыми ресурсами

6.2.4.1 Цель

Целью процесса управления кадровыми ресурсами является обеспечение организации необходимыми кадровыми ресурсами и поддержание их компетенций в соответствии с деловыми потребностями.

Данный процесс обеспечивает наличие компетентного и опытного персонала, достаточно квалифицированного для выполнения процессов жизненного цикла, направленных на достижение целей организации, проекта и заинтересованных сторон.

6.2.4.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.2.4 *[1]* и 6.2.4 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.2.4.3 Особенности, специфические для ИИ

Для процесса управления кадровыми ресурсами дополнительные действия или задачи не определены.

Использование методов ИИ вовлекает в жизненный цикл исполнителей новых ролей. Например, специалисты по интеллектуальному анализу данных (data scientists), специалисты по инженерии данных (data engineers) играют дополнительные роли в качестве разработчиков ИИ в области машинного обучения. Инженеры по знаниям играют дополнительную роль в качестве разработчиков ИИ в инженерии знаний. При выполнении действий и задач, указанных в 6.2.4.2, организации должны принимать во внимание навыки и компетенции исполнителей этих дополнительных ролей.

Кроме того, организациям, только начинающим осваивать ИИ, следует проанализировать имеющиеся кадровые ресурсы и оценить адекватность их компетенций.

Более подробную информацию о ролях заинтересованных сторон (таких, как разработчики ИИ, поставщики ИИ, поставщики данных) см. в *ГОСТ Р 71476—2024* (пункт 5.17).

6.2.5 Процесс управления качеством

6.2.5.1 Цель

Целью процесса управления качеством является обеспечение соответствия продуктов, услуг и внедрений (реализаций) целям организаций и проектов с точки зрения качества, а также удовлетворение соответствующих требований потребителей.

6.2.5.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.2.5 *[1]* и 6.2.5 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.2.5.3 Особенности, специфические для ИИ

Для процесса управления качеством дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.2.5.2, организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ.

Организации следует предусмотреть реализацию специфических для ИИ особенностей, описанных в процессах управления качеством, путем разработки соответствующих политик, целей, процедур и т. п. Обеспечение качества и его оценка могут играть более заметную роль в организациях, разрабатывающих, развертывающих системы ИИ и ведущих их мониторинг.

Действия по непрерывному управлению качеством поддерживают систематическую оценку показателей работы системы ИИ на протяжении всего ее жизненного цикла, включая поддержание единого уровня качества с момента ее развертывания.

6.2.6 Процесс управления знаниями

6.2.6.1 Цель

Целью процесса управления знаниями являются развитие способностей (потенциала) и создание активов, позволяющих организации воспользоваться возможностями для повторного использования имеющихся знаний.

Данный процесс охватывает знания, навыки и компетенции, а также активы знаний, включающие элементы систем.

Знания, которые используются для создания моделей ИИ, обсуждаются в 6.4.7 и 6.4.9.

6.2.6.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.2.6 *[1]* и 6.2.6 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.2.6.3 Особенности, специфические для ИИ

Для процесса управления знаниями дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.2.6.2, организациям следует принять во внимание следующие специфические для ИИ особенности:

- следует подумать об охвате управлением знаниями элементов систем ИИ (таких, например, как наборы данных и сценарии подготовки данных) наравне с любыми другими элементами систем;
- экспериментирование является важным аспектом в реализации систем ИИ. Документирование экспериментов играет важную роль в предотвращении повторения в будущем ранее проведенных экспериментов как той же, так и иной заинтересованной стороной. Кроме того, материалы, документирующие эксперименты, содержат важные знания и опыт, которые можно использовать для дальнейших улучшений;
- более подробные сведения об управлении кадровыми ресурсами с точки зрения опыта интеллектуального анализа данных см. в 6.2.4;
- более подробные сведения, касающиеся происхождения и дальнейшей истории обработки и хранения данных, см. в 6.4.8.

6.3 Процессы технического управления

6.3.1 Процесс планирования проекта

6.3.1.1 Цель

Целью процесса планирования проекта является создание и координация эффективных и выполнимых планов.

Данный процесс:

- определяет диапазон управления проектом и технической деятельности;
- определяет результаты процесса, задачи и итоговые материалы;
- устанавливает графики выполнения задач, включая критерии достижения целей;
- оценивает необходимые для выполнения задач ресурсы.

Процесс планирования является непрекращающимся процессом, который продолжается на протяжении всего проекта, при этом регулярно производится пересмотр планов.

6.3.1.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.1 *[1]* и 6.3.1 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.3.1.3 Особенности, специфические для ИИ

Для процесса планирования проекта дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.1.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ.

При реализации вида деятельности «планирование проекта и техническое управление» важно предусмотреть определенную гибкость в отношении создания модели (см. 6.3.1.3 *[1]* и 6.3.1.3 *[2]*). Обес-

печение предсказуемости разработки программного обеспечения уже является сложной задачей, и это в еще большей степени справедливо в отношении предсказуемости создания модели. Создание модели может потребовать инженерии данных для ИИ, такой как сбор данных, разметка (аннотация) данных и предварительная обработка данных (см. 6.4.8). Для системы ИИ на основе машинного обучения создание модели может потребовать итераций экспериментирования и проведения экспериментов с использованием различных стратегий и тактик достижения желаемых производительности и качества модели. Для системы ИИ на основе инженерии знаний создание модели может включать приобретение знаний и извлечение знаний (knowledge elicitation).

Кроме того, при планировании проекта следует учитывать различные другие специфические для ИИ особенности вовлеченных процессов, например: организация непрерывной валидации (см. 6.4.14).

6.3.2 Процесс оценки и контроля проекта

6.3.2.1 Цель

Целями процесса оценки и контроля проекта являются:

- оценка согласованности и выполнимости планов;
- определение статуса (состояния) проекта, технических показателей и показателей процессов;
- корректировка хода исполнения проектов, помогающая обеспечить выполнение проектов в соответствии с планами, графиками, запланированными бюджетами и поставленными техническими задачами.

В данном процессе как периодически, так и при наступлении всех основных событий оцениваются прогресс и достижения проекта в сравнении с требованиями, планами и общими деловыми целями. В случае выявления существенных отклонений информация предоставляется руководству для принятия соответствующих мер. Данный процесс может включать переориентацию видов деятельности и задач проекта с целью корректировки изменений и отклонений от других процессов технического управления и технических процессов. Переориентация допускает перепланирование.

6.3.2.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.2 *[1]* и 6.3.2 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.3.2.3 Особенности, специфические для ИИ

Для процесса оценки и контроля проекта дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.2.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ.

При реализации действия «планирование оценки проекта и контроля над ним» могут быть установлены интервалы времени (как определено в процессе управления качеством) для обновления системы и/или модели ИИ (см. 6.3.2.3 *[1]* и 6.3.2.3 *[2]*).

Ход реализации системы ИИ менее предсказуем из-за ее итеративного и экспериментального характера. Например, прогресс в данном случае невозможно надежно измерить, подсчитав количество написанных строк кода.

6.3.3 Процесс управления решениями

6.3.3.1 Цель

Цель процесса управления решениями заключается в том, чтобы обеспечить структурированную, аналитическую концептуальную основу для объективного выявления, характеризации и оценивания ряда альтернативных решений в любой момент жизненного цикла и для выбора оптимальной последовательности действий.

6.3.3.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.3 *[1]* и 6.3.3 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.3.3.3 Особенности, специфические для ИИ

Для процесса управления решениями дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.3.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ.

Использование ИИ добавляет системе неопределенности и сложности за счет введения новых типов решений (более подробно о качестве систем ИИ и моделей машинного обучения см. в *[14]* и *[15]*).

Новые типы решений включают (не ограничиваясь ими):

- решения об отказе от использования системы ИИ, если результаты ее функционирования более не соответствуют требованиям;

- решение о «реорганизации» системы ИИ в случае, когда обучение модели приводит к тому, что результаты ее функционирования более не соответствуют требованиям (т. е. система перезапускается, и создается новая обученная модель);

- решение об обновлении спецификаций и контрактов между приобретающей стороной, пользователем и/или поставщиком с целью отразить в них приобретенное в результате обучения поведение;

- решение об обновлении документации с целью отразить в ней приобретенное в результате обучения поведение (при условии, что система в ходе функционирования продолжает соответствовать требованиям и контракту).

Например, организация должна определить, как измеряется качество моделей МО при реализации действия «Анализ информации о решениях» (см. 6.3.3.3 *[1]* и 6.3.3.3 *[2]*). Такие аспекты усиливают важность заранее определенного процесса принятия решений.

Примечание — После того как организация приняла решение об использовании ИИ, процесс управления решениями может помочь руководящим органам определить моменты принятия решений, в которых могут возникнуть и могут быть решены ключевые вопросы стратегического управления (см. 5.3 *[16]*).

6.3.4 Процесс управления рисками

6.3.4.1 Цель

Процесс управления рисками представляет собой непрекращающийся процесс, цель которого заключается в непрерывном систематическом выявлении, анализе, обработке и мониторинге рисков на протяжении всего жизненного цикла системы, продукта или услуги. Его можно применять в отношении рисков, связанных с приобретением, разработкой, сопровождением и функционированием системы.

6.3.4.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.4 *[1]* и 6.3.4 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.3.4.3 Особенности, специфические для ИИ

Для процесса управления рисками дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.4.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ, а также обратиться к *[7]* по поводу деталей управления рисками систем ИИ. Определенные в *[7]* цели управления рисками включают справедливость, неприкосновенность частной жизни, надежность, прозрачность, объяснимость, подотчетность, доступность, целостность и сопровождаемость.

Действия в рамках процесса оценки рисков должны охватывать все риски, связанные с системой ИИ, и включать реализацию адекватных мер по обработке рисков посредством использования плана обработки рисков и соответствующих документов, относящихся к управлению рисками и описанных в *[7]*. Стандарт *[7]* содержит рекомендации по управлению рисками для организаций, которые разрабатывают, производят, развертывают и применяют продукты, системы и услуги, использующие ИИ. Он не предназначен для управления рисками организациями, которые используют такие продукты и услуги для обеспечения безопасности и защищенности. Таким образом, организациям, которые применяют ИИ в продуктах и услугах, нацеленных на обеспечение безопасности и защищенности, следует принять во внимание применимые международные стандарты управления рисками в дополнение к учету специфических для ИИ особенностей в отношении процессов в соответствии с *[7]*. Соображения по вопросу обеспечения функциональной безопасности систем ИИ можно найти в техническом отчете *[3]*. Так, например, разработчики систем ИИ, рассматриваемых как медицинские устройства, должны обеспечивать управление рисками в соответствии с такими международными стандартами, как *[17]*.

В дополнение к рекомендациям, содержащимся в 6.3.4 *[1]* и 6.3.4 *[2]*, у систем ИИ по сравнению с традиционными программными системами имеются дополнительные области возможностей и проблемные зоны. Такие области отмечены и более подробно объяснены в *[7]*.

Еще одно специфическое для ИИ соображение применимо в том случае, когда системы ИИ запрограммированы на вычисление в автономном режиме оперативных решений, связанных с риском причинения вреда, и когда такие решения из-за ограничений по времени не могут быть проверены человеком (примером могут послужить некоторые решения, принимаемые беспилотными транспортными средствами). Такие риски можно смягчить посредством постоянного управления рисками самой системой. Простой формой этого является установление определенных границ, в рамках которых система может работать. Например, автоматизированная система управления микроклиматом не должна допускать нагрева до опасного диапазона температур. Определенные правила могут быть использованы для управления рисками — например, запрет автоматического открытия багажника на высокой скорости

даже при команде водителя. Наиболее продвинутый тип непрерывного управления рисками — это когда система ИИ активно выполняет анализ связанных с решениями рисков посредством логических рассуждений, основанных на модели мира и правилах. Помимо автономного управления рисками, риск причинения вреда может быть смягчен за счет достаточного покрытия тестовыми вариантами, позволяющего убедиться в том, что в рискованных ситуациях не будут приниматься решения, влекущие за собой причинение вреда. Кроме того, в отношении машинного обучения следует также уделить особое внимание включению в качестве примеров в обучающие данные чреватых рисками ситуаций и соответствующих правильных решений. Также на более позднем этапе человеком в целях обеспечения качества может быть выполнен ретроспективный анализ автоматизированного управления рисками.

О процессе определения системных требований в отношении важных свойств систем ИИ см. также 6.4.3. В дополнение к типичным рискам, рассматриваемым для системы, таким как риски для безопасности и неприкосновенности частной жизни (защиты персональных данных), план обработки рисков должен также включать риски, связанные с установленными организацией целями.

Организациям следует выявлять потенциальные риски и возможности, связанные с системами ИИ, в том числе проводить консультации с типичными пользователями и иными заинтересованными сторонами для выяснения их потребностей и требований (6.4.2).

Дополнительные требования к управлению рисками могут быть применимы в зависимости от назначения системы ИИ и от нормативно-правовой среды, в рамках которой предполагается использовать систему ИИ.

Если система ИИ имеет отношение к безопасности, то для обеспечения подотчетности организация должна иметь журналы аудита (audit trail), включающие сведения о происхождении данных, о валидации источника данных, об анализе и смягчении рисков, а также о решениях. Такой подход может быть также рекомендован и для других систем ИИ. Таким образом, разработка системы ИИ должна включать разработку стратегии ее аудита. Примером может служить сохранение ранее принятых решений вместе со ссылкой на использованную модель, включая сведения о том, как эта модель была создана. Стратегия аудита может включать в себя документирование ключевых решений, принятых в ходе самого процесса разработки, и их обоснований (например, почему было отдано предпочтение определенной модели).

6.3.5 Процесс управления конфигурацией

6.3.5.1 Цель

Целью процесса управления конфигурацией являются управление и контроль над элементами и конфигурацией системы на протяжении всего ее жизненного цикла. Управление конфигурацией также обеспечивает согласованность между продуктом и связанным с ним определением конфигурации.

Примечание — Более подробная информация об управлении конфигурацией приведена в *[18]*.

6.3.5.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.5 *[1]* и 6.3.5 *[2]*, касающиеся выполняемых действий, задач и результатов процесса, со следующим дополнением.

В соответствии с применимыми политиками и процедурами организации в отношении процесса управления конфигурацией в рамках проекта необходимо реализовать следующее действие: автоматизированный процесс отката модели может быть использован для быстрого устранения неоптимальной производительности модели.

6.3.5.3 Особенности, специфические для ИИ

Для процесса управления конфигурацией дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.5.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ.

Помимо традиционных компонентов программного обеспечения и конфигурации, системы ИИ содержат специфические для ИИ артефакты, в отношении которых также требуется управление конфигурацией: представляющие модель данные (например, правила, весовые коэффициенты, параметры), документация на элементы ИИ, данные и метаданные. В случае использования машинного обучения может быть полезно применить управление конфигурацией в отношении модели в сочетании с данными, на которых она была обучена. Это обеспечит отслеживаемость (например, для целей аудита и исполнения нормативно-правовых требований) и воспроизводимость экспериментов.

По сравнению с артефактами традиционного программного обеспечения (такими, например, как исходный код, тестовые примеры, тестовые данные) артефакты систем ИИ — особенно наборы данных — могут быть большими по объему и обычно хранятся в системах отдельно от программного кода и

конфигурационных файлов. В некоторых системах ИИ репозиторий (хранилище) более старых данных следует сохранять в нетронутом виде на случай возможной необходимости отката версии приложения. Это может привести к выбору решений, отходящих от типичной для традиционного программного обеспечения практики, — например, к установлению более коротких сроков хранения для версий.

Типичным применением управления конфигурацией для системы ИИ является откат к предыдущей версии модели среды выполнения, когда у новой модели выявляются проблемы с качеством. Обрабатываемая в процессе управления конфигурацией информация включает данные, которые используются для построения и тестирования модели ИИ. Более подробную информацию об этих данных можно найти в 6.4.7 и 6.4.8. Кроме того, использовавшиеся для построения и тестирования модели ИИ данные также охватываются управлением конфигурацией, в рамках которого работоспособность системы ИИ непрерывно контролируется и поддерживается (см. 6.4.14, 6.4.15 и 6.4.16).

Кроме того, управления версиями системы ИИ оказывается уже недостаточно для получения четкого представления о конфигурации, поскольку версии элементов конфигурации для целей разработки и логистики более не отражают гарантированное поведение, ассоциируемое с рабочей конфигурацией. В частности, если развернуто несколько экземпляров в одной и той же конфигурации, их поведение может различаться.

На стадиях проектирования и разработки организации следует подумать об использовании специфических для ИИ средств управления исходным кодом, учитывающих характерные для ИИ особенности (например, инженерию данных для ИИ, обучение моделей).

6.3.6 Процесс управления информацией

6.3.6.1 Цель

Цель процесса управления информацией заключается в том, чтобы для (или в интересах) обозначенных заинтересованных сторон производить, получать, подтверждать, преобразовывать, сохранять, извлекать, распространять и уничтожать информацию либо передавать ее на архивное хранение.

В рамках процесса управления информацией осуществляются планирование, приведение планов в исполнение и контроль над предоставлением обозначенным заинтересованным сторонам однозначной, полной, проверяемой, непротиворечивой, отслеживаемой информации, представленной в допускающей модификацию форме и в удобном для восприятия виде. В состав информации входит техническая, проектная, организационная, договорная и пользовательская информация. Информация часто извлекается из записей данных организации, системы, процесса или проекта.

6.3.6.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.6 *[1]* и 6.3.6 *[2],* касающиеся выполняемых действий, задач и результатов процесса.

6.3.6.3 Особенности, специфические для ИИ

Для процесса управления информацией дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.6.2, проектным командам и/или организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ.

Системы ИИ, как правило, очень интенсивно обрабатывают данные и используют наборы данных для тестирования и, в случае машинного обучения, также для обучения. Эти наборы данных являются частью информации, которой следует управлять (см. 6.4.8).

6.3.7 Процесс измерений

Применимы положения 6.3.7 *[1]* и 6.3.7 *[2],* касающиеся цели, выполняемых действий, задач и результатов процесса.

Кроме того, если система ИИ имеет отношение к безопасности, то необходимо рассмотреть процессы специфических для ИИ измерений (например, вероятности получения ошибочного результата); такие же процессы рекомендуются и для других систем ИИ. В частности, с целью проведения корректировок может быть измерен дрейф концепции и/или данных в моделях ИИ, вызванный как изменениями во внешних условиях, так и изменениями в самой системе.

6.3.8 Процесс обеспечения качества

6.3.8.1 Цель

Целью процесса обеспечения качества является помощь в эффективном применении процессов управления качеством в организации в рамках конкретного проекта.

В центре процесса обеспечения качества находится обеспечение уверенности в том, что требования к качеству выполнены. Проводится упреждающий анализ процессов жизненного цикла проекта и их результатов с целью обеспечить желаемое качество разрабатываемого продукта, а также соблюдение политик и процедур организации и проекта.

6.3.8.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.3.8 *[1]* и 6.3.8 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

Обеспечение качества как часть процесса управления качеством и его оценка могут играть более заметную роль в организациях, разрабатывающих, развертывающих и ведущих мониторинг систем ИИ.

6.3.8.3 Особенности, специфические для ИИ

Для процесса обеспечения качества дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.3.8.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, специфические для ИИ.

В дополнение к рекомендациям, содержащимся в 6.3.8 *[1]* и 6.3.8 *[2]*, обеспечение качества может играть более заметную роль для систем ИИ по сравнению с традиционными программными системами. Системы ИИ способны эволюционировать с течением времени (например, в случае систем непрерывного обучения). Такая эволюция делает необходимыми усилия по тщательному мониторингу и обеспечению качества с целью выявления возможного падения эффективности, вызванного, например, низким качеством данных на входе в модель, дрейфом концепции или дрейфом данных.

В рамках процесса обеспечения качества проводятся мониторинг и оценка как продукта, так и процесса. В системах ИИ алгоритмы и данные для машинного обучения также рассматриваются как подлежащие оценке продукты. При оценке этих продуктов следует дополнительно рассматривать показатели качества, специфические для систем ИИ (такие, как, например, прозрачность, справедливость, подотчетность, устойчивость к изменениям). Дополнительная информация об аспектах качества систем ИИ приведена в *[14]* и *[15]*.

Кроме того, подлежащие оценке процессы должны включать действия по проведению анализа во время апробации концепции; задачи по проведению анализа требований и рисков с целью обеспечения надлежащего охвата интересующей проблемной области; итерационные задачи по проведению машинного обучения и/или процедуры по созданию обучающих данных (сбор, отбор, генерация, валидация, модификация или добавление).

Более подробная информация об этих данных, процессах и об оценке их качества в контексте машинного обучения приведена в 6.4.7, 6.4.8 и 6.4.14.

В качестве примеров событий, мониторинг которых следует проводить в рамках обеспечения качества, можно назвать следующие:

- подаваемые на вход модели данные имеют низкое качество;
- оцениваемые моделью данные подвержены изменениям (дрейф данных);
- наблюдаются отклонения от желаемого результата (дрейф концепции).

Действия по обеспечению качества должны соответствовать характеру использования системы ИИ. Как правило, на уровень, на котором организациям следует осуществлять деятельность по обеспечению качества, оказывают влияние сложность среды, уровень автономии, а также воздействие результатов работы системы ИИ. Кроме того, могут существовать внешние факторы, такие как нормативные правовые требования и требования систем менеджмента качества, которые влияют на тип и масштабы деятельности по обеспечению качества. Организациям следует — в особенности в отношении систем ИИ с непрерывным обучением — подумать о выполнении соответствующих действий по повторной валидации.

Подробности см. в 6.2.5, 6.4.11, 6.4.13, 6.4.14 и в описании анализа качества данных в 6.4.8.

6.4 Технические процессы

6.4.1 Процесс анализа миссии и/или деятельности

6.4.1.1 Цель

Цель процесса анализа миссии и/или деятельности заключается в том, чтобы определить проблемы и возможности, связанные с выполнением миссии и/или ведением дел, охарактеризовать пространство решений и определить потенциальный класс или классы решений, которые позволят преодолеть проблемы или воспользоваться возможностями.

6.4.1.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.1 *[1]* и 6.4.1 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.1.3 Особенности, специфические для ИИ

Для процесса анализа миссии и/или деятельности дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.4.1.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ.

Использование систем ИИ сопряжено со специфическими рисками (см. 6.3.4), которые могут повлиять на достижение определенных деловых целей или даже сделать его невозможным. Например, при выполнении действия «Определение областей возможностей и проблемных зон» организация обязана учитывать, что требования законодательства о защите персональных данных могут исключить возможность использования персональных данных для целей, отличающихся от целей обработки, первоначально указанных при их сборе (см. 6.4.1.3 *[1]* и 6.4.1.3 *[2]*). В случае принятия решений, способных негативно повлиять на физических лиц, законодательство может потребовать объяснения того, какие данные были использованы и как именно.

Другими примерами рисков являются степень доступности данных и качество данных.

6.4.2 Процесс определения потребностей и требований заинтересованных сторон

6.4.2.1 Цель

Цель процесса определения потребностей и требований заинтересованных сторон заключается в том, чтобы выявить и зафиксировать требования заинтересованных сторон к системе с тем, чтобы система могла предоставлять возможности, необходимые пользователям и иным заинтересованным сторонам в заданной среде применения.

В ходе данного процесса выявляются заинтересованные стороны и/или категории заинтересованных сторон, а также их потребности на протяжении всего жизненного цикла системы ИИ. Эти потребности анализируются и трансформируются в единый набор требований заинтересованных сторон, отражающий желаемое взаимодействие системы со средой ее эксплуатации и являющийся базовым документом, в сопоставлении с которым проводится валидация каждой из разработанных функциональных возможностей. Требования заинтересованных сторон определяются с учетом контекста взаимодействия рассматриваемой системы с другими системами (включая обеспечивающие).

6.4.2.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.2 *[1]* и 6.4.2 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.2.3 Особенности, специфические для ИИ

Для процесса определения потребностей и требований заинтересованных сторон дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.4.2.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ.

Ввиду индуктивного характера систем ИИ крайне важно выполнить задачу «получение явного согласия в отношении требований заинтересованных сторон», включая критически важные показатели производительности, которые дают возможность оценивать в качестве цели технические достижения (см. 6.4.2.3 *[1]* и 6.4.2.3 *[2]*). Организациям следует принять во внимание риск проявления необъективности и предвзятости вследствие узости взглядов заинтересованных сторон.

Такое техническое достижение должно быть специфицировано организацией для того, чтобы сделать возможным мониторинг целей посредством процесса обеспечения уверенности в качестве (см. 6.3.8).

Характер использования системы ИИ может служить для выделения отдельных типов заинтересованных сторон, которые следует принять во внимание. В их число входят:

- поставщики платформ, продуктов или услуг ИИ;
- разработчики ИИ;
- заказчики и пользователи;
- партнеры, занимающиеся системной интеграцией, предоставлением данных и аудитом;
- определяющие политику и регулирующие органы, субъекты данных;
- другие лица, которых затрагивает разработка и использование системы ИИ.

В процессе выявления заинтересованных сторон могут быть получены данные, которые будут направлять разработку элементов системы, включая пользовательский интерфейс, документацию и варианты использования. Организациям следует дополнительно более глубоко изучить и уточнить эти сведения, до такой степени, чтобы они могли стать частью системных требований. В уточнении этих данных также могут помочь структуры, описывающие нормативное правовое регулирование, права человека, социальную ответственность, экологические проблемы. Для получения более подробной ин-

формации о возможных типах заинтересованных сторон применительно к ИИ см. *ГОСТ Р 71476—2024* (пункт 5.17).

Примечание — Описанные в *[14]* показатели качества, входящие в состав модели качества систем ИИ, полезны для выявления и идентификации требований к качеству среди нефункциональных требований, которые часто представляют собой неявно выраженные потребности заинтересованных сторон. Для получения дополнительной информации об оценке качества систем ИИ см. также *[15]*.

6.4.3 Процесс определения системных требований

6.4.3.1 Цель

Целью процесса определения системных требований является изучение всех требований заинтересованных сторон и их трансформация в техническое видение решения, которое по-прежнему будет отвечать эксплуатационным потребностям пользователя. В частности, данный процесс принимает во внимание результаты процессов управления рисками и стратегического управления, как показано на рисунке 2.

Данный процесс создает набор измеримых системных требований, задающих для поставщика (выполняющего роль производителя ИИ, партнера ИИ или поставщика ИИ) характеристики, атрибуты, функциональные и эксплуатационные возможности, которыми система должна обладать для удовлетворения требований заинтересованных сторон. Насколько это позволяют имеющиеся ограничения, эти требования не должны подразумевать какую-либо конкретную реализацию.

6.4.3.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.3 *[1]* и 6.4.3 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.3.3 Особенности, специфические для ИИ

Для процесса определения системных требований дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.4.3.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ:

- желаемые эксплуатационные показатели (степень корректности) модели или моделей. Установление этих требований диктует необходимость тщательного выбора правильных метрик (например, минимальной точности и минимальной прецизионности). Такие требования могут включать в себя диапазон входных данных, для которых модель должна функционировать в требуемых границах. Например, модель в 90 % случаев должна быть способна отличить кошку от собаки на сделанных в дневное время фотографиях, на которых животное видно целиком;
- требования к степени автономности системы ИИ. К ним относятся соображения, касающиеся реализуемого системой ИИ уровня автономности — например, присутствует ли человек в контуре управления. Если да, то устанавливается, какие решения человек может принимать в отношении поведения системы ИИ (такие, например, как установка или корректировка пороговых значений, настраивающих желаемый уровень функционирования системы ИИ);
- требования к тому, как следует реагировать в случае непредвиденного поведения системы, — например, путем установления и применения дополнительных детерминированных правил с целью обеспечения безопасности;
- требования к производительности системы: следует установить такие требования, как, например, желаемое время выполнения, которое часто зависит от типа используемой модели;
- требования к прозрачности и объяснимости: модели машинного обучения могут быть очень сложными и, как следствие, трудными для понимания. В зависимости от ситуации у физических лиц может иметься законное право требовать объяснений того, как моделью было принято решение, особенно в том случае, если они при этом серьезно пострадали (например, в правовом или финансовом отношении). Например, в некоторых случаях требуется давать объяснения в случае отказа в предоставлении кредита. Характер объяснений может варьироваться от детального до высокоуровневого описания того, какие данные и какой тип алгоритма машинного обучения использовались. Объяснения могут помочь с обеспечением приемлемости решений ИИ, но они также могут привести к проблемам в тех случаях, когда объяснение указывает на наличие ошибки;
- предполагается, что организация в соответствии с применимыми нормативно-правовыми требованиями информирует физических лиц о том, что они взаимодействуют с системой ИИ;
- требования к непрерывной валидации: см. описание процесса непрерывной валидации (см. 6.4.14);

- требования к обеспечению справедливости: важно установить требования к обеспечению справедливости и инклюзивности алгоритма и данных в отношении определенных групп в обществе. Кроме того, решения системы ИИ должны основываться на четких и понятных характеристиках с тем, чтобы можно было проверить их справедливость. Чтобы установить такие требования, следует определить метрики справедливости;

- требования к защите неприкосновенности частной жизни (персональных данных): применимы в случае обработки персональных данных. Важное значение имеют информирование физических лиц, предоставление им возможности контроля и обеспечение защиты персональных данных. Кроме того, соображения, связанные с защитой персональных данных, могут повлиять на выбор алгоритма (например, могут быть использованы алгоритмы дифференциальной защиты персональных данных (differential privacy), см. *[19]*);

- требования безопасности: применимы в случае, если в результате использования ИИ появится дополнительная поверхность атаки. Обычно в их число входят следующие:
  - обеспечение защиты данных, используемых либо для обучения, либо для тестирования, либо для того и другого вместе, включая защиту от атак «отравления или порчи данных», когда злоумышленники вбрасывают данные, чтобы повлиять на поведение моделей машинного обучения;
  - обеспечение защиты от манипулирования входными данными (когда, например, нежелательные сообщения электронной почты (спам) классифицируются как «не спам»);
  - обеспечение защиты от «инверсии модели» — ситуации, в которой злоумышленнику удается путем деконструирования извлечь чувствительные данные, использованные для обучения модели;
  - обеспечение защиты от «кражи модели», когда злоумышленник пытается скопировать поведение модели, являющейся интеллектуальной собственностью.

6.4.4 Процесс определения архитектуры системы

Применимы положения 6.4.4 *[1]* и 6.4.4 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

Примечание — Для процесса определения архитектуры системы в 6.4.4 *[1]* используется название «процесс определения архитектуры».

6.4.5 Процесс выбора проектно-конструкторских решений

Применимы положения 6.4.5 *[1]* и 6.4.5 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

6.4.6 Процесс системного анализа

Применимы положения 6.4.6 *[1]* и 6.4.6 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

6.4.7 Процесс приобретения знаний

6.4.7.1 Цель

Целью процесса приобретения знаний является получение знаний, необходимых для создания моделей ИИ.

Для многих систем ИИ знания о предметной области и о проблеме играют первостепенную роль.

В системах ИИ на основе машинного обучения знания используются для того, чтобы направлять ход выполнения задач отбора данных, подготовки данных и разработки моделей. Процесс приобретения знаний может осуществляться путем проведения исследований и/или посредством привлечения экспертов в предметной области.

В случае основанной на знаниях системы ИИ знания должны быть явным образом закодированы в модели.

Анализ данных (см. 6.4.8) может играть роль в сборе и уточнении знаний.

Примечание — Знания, о которых идет речь в описании процесса приобретения знаний, — это знания, необходимые для создания моделей ИИ.

6.4.7.2 Результаты процесса

В результате успешного выполнения процесса приобретения знаний должны быть достигнуты следующие результаты:

a) выявлены знания, необходимые для создания моделей ИИ;

b) сохранены собранные знания;

c) обеспечена прослеживаемость в процессе приобретения знаний.

6.4.7.3 Действия и задачи

В рамках проекта в соответствии с применимыми политиками и процедурами организации в отношении процесса приобретения знаний необходимо реализовать следующие действия:

a) определение сферы охвата и критериев для приобретения знаний. В качестве первого шага определяются сфера охвата и критерии для приобретения знаний: к какой предметной области и какому аспекту относятся знания? Насколько актуальны эти знания?

b) поиск и подбор источников знаний. Знания могут быть извлечены из публикаций и данных либо получены от экспертов;

c) приобретение знаний с целью извлечения знаний. Для того чтобы воспользоваться знаниями, можно изучать публикации, анализировать данные, проводить собеседования с экспертами или наблюдать за ними. В случае инженерии знаний извлеченные знания должны быть формализованы таким образом, чтобы задействованные алгоритмы могли их использовать. Эти действия являются частью процесса реализации (см. п. 6.4.9);

d) сбор знаний о предметной области и проблеме посредством изучения, проведения собеседований или использования иных способов извлечения знаний, анализа данных, приобретения документированных знаний и/или привлечения заинтересованных сторон, располагающих необходимыми знаниями;

e) управление результатами приобретения знаний.

Примечания

1 Роли, действия, структурные уровни, компоненты инженерии знаний и их взаимосвязи, а также общеупотребительная терминология инженерии знаний представлены в *[20]*.

2 Коллективное использование несколькими командами проектов собранных для каждого проекта знаний может осуществляться с помощью репозиториев (областей, в которых хранятся наборы знаний) и реестров (систем или средств регистрации использования наборов знаний). В ходе процесса сбора знаний может быть рассмотрена возможность повторного использования знаний, касающихся апробированных типовых решений, которые могут быть применимы в деятельности по разработке модели (см. 6.4.9.3).

6.4.8 Процесс инженерии данных для ИИ

6.4.8.1 Цель

Цель процесса инженерии данных для ИИ заключается в обеспечении возможности использования данных для создания моделей ИИ и их верификации. Данные занимают центральное место в инженерии моделей машинного обучения, поскольку они используются для их обучения. Для эвристических моделей роль данных при создании модели более вторична, поскольку в этом случае они могут использоваться для поддержки инженерии знаний (см. 6.4.9).

6.4.8.2 Результаты процесса

В результате успешного выполнения процесса инженерии данных для ИИ должны быть достигнуты следующие результаты:

a) выявлены требуемые данные и наборы данных, проведен анализ выборок из них и организовано их получение;

b) обучающие данные и, при необходимости, валидационные (проверочные) данные подготовлены, отформатированы и сделаны доступными для моделей машинного обучения;

c) подготовлены тестовые данные для валидации (см. 6.4.13) и/или верификации (см. 6.4.11);

d) подготовлены данные для ручного анализа, проводимого ради достижения более глубокого понимания с целью поддержки процессов инженерии данных для ИИ и инженерии моделей;

e) выявлены автоматизированные процессы (если таковые имеются) для извлечения, преобразования и загрузки данных;

f) любая запись и любое использование персональных данных в составе данных соответствуют применимым законам и нормативно-правовым требованиям;

g) подготовлены артефакты (такие, как метаданные) для отслеживания, документирования и поддержки данных и автоматизированных процессов, включая процесс управления конфигурацией;

h) данные своевременно удаляются;

i) обеспечено управление мультимодальными данными.

Примечание — Поскольку мультимодальный тип данных (например, речь, изображения, данные с воспринимающих устройств) все чаще встречается в системах ИИ, то могут быть использованы наилучшие практики для обработки, проектирования и развертывания мультимодальных (с комбинированным вводом данных) систем ИИ.

6.4.8.3 Действия и задачи

В рамках проекта должны быть реализованы следующие действия в соответствии с политиками и процедурами организации, применимыми в отношении процесса инженерии данных для ИИ.

a) Приобретение и/или отбор данных

Целью модели ИИ является создание выходных данных на основе входных данных (например, классификация животного на основе поданного на вход изображения), поэтому данные следует собирать для формирования таких комбинаций входных и выходных данных. Типичными формами данных являются структурированные данные, текст, звук, изображения и иные данные, поступающие от воспринимающих устройств (сенсоров).

Примерами способов сбора данных являются:

- сбор данных из существующего хранилища данных (например, данные о клиентах);
- запись данных о ходе процесса (например, с промышленных датчиков);
- запись данных о ходе срежиссированного процесса (например, отыгранных актерами сцен с целью создания видеопримеров обнаружения определенных событий).

Для проверки способности модели машинного обучения обобщать за рамками обучающего набора данных полезно, чтобы тестовые данные происходили из другого источника или процесса. Известным примером является случай, когда модель машинного обучения научилась распознавать волков на основе размеченных обучающих данных. Оказалось, что модель могла хорошо работать потому, что все обучающие фотографии волков были сделаны зимой и их легко можно было идентифицировать по наличию снега. Чтобы избежать подобных проблем обобщения, тестовые данные следовало собирать из иного источника.

Процесс приобретения и отбора данных должен быть непрекращающимся или регулярно повторяющимся в ситуациях, когда взаимосвязь между входными и выходными переменными со временем меняется. Например, чтобы спрогнозировать цену продажи участка земли, важно быть в курсе изменений в экономике и на рынке, которые отражаются в новых данных. Эти новые данные могут быть использованы для регулярного тестирования модели и, при необходимости, для ее переобучения или перепроектирования (см. 6.4.14). Более старые данные, отражающие устаревшие взаимосвязи, по тем же причинам следует удалять.

b) Выполнение разметки (аннотирования) данных

Разметка данных представляет собой особую форму приобретения данных, когда образцам присваиваются значения желаемых результатов их классификации — например, изображения животных помечаются словами «кошка» или «собака». Обычно это делается вручную, поэтому реализация строго контролируемого процесса может помочь предотвратить появление предвзятости или шума вследствие субъективности.

Выполняющие разметку данных лица должны быть компетентны в области, к которой относится разметка, и обучены использованию инструмента разметки. В зависимости от степени риска, связанной с приложением, результаты разметки могут быть пересмотрены и при необходимости скорректированы.

При использовании инструментов, помогающих проводить разметку данных, организациям следует оценить статус используемых для процесса разметки инструментов. Такая оценка должна включать оценку особенностей и функциональных возможностей подобных средств аннотирования и надлежащую валидацию этих инструментов с целью обеспечения высокого качества размеченных данных (см. 6.4.8.3, e) и f)).

c) Анализ и изучение данных с целью их понимания

Собранные данные могут быть проанализированы и изучены, что поможет понять предметную область, проблему и связанные с данными вопросы. Для машинного обучения такое понимание может привести к новым идеям и представлениям о том, какие иные данные необходимы и/или какая требуется обработка данных. Для инженерии знаний анализ данных может быть полезен для дальнейшей организации и систематизации существующих знаний.

d) Анализ качества данных

У данных может иметься много проблем с качеством, требующих проведения оценки с целью управления проведением выбора, очистки и корректировки данных. Данные должны иметься в достаточном количестве, а погрешность должна быть в допустимых пределах. Данные должны быть достаточно полными (неоднородными и разнообразными), с тем чтобы адекватно представлять ожидаемые эксплуатационные данные. Обучающие данные предпочтительно должны иметь тот же баланс (распределение), с которым, как ожидается, столкнется модель, однако при этом необходимо учесть специфические пограничные ситуации.

Наличие предвзятости (которая может возникнуть, например, вследствие субъективных решений) можно проконтролировать, проверив, хорошо ли сбалансировано желаемое поведение по отношению к социальным признакам, дискриминация по которым запрещена (таким, как пол или этническая принадлежность). Удаления таких признаков часто недостаточно для устранения предвзятости, поскольку входные данные могут по-прежнему содержать элементы данных, которые фактически эквивалентны этим атрибутам.

Особым аспектом качества данных является риск отравления данных: злоумышленник может изменить, удалить или добавить данные с целью нежелательным образом повлиять на поведение модели.

Анализ качества данных, как правило, представляет собой непрекращающийся процесс, поскольку со временем могут возникать новые проблемы с качеством, поэтому рекомендуется автоматизировать проверки и верификацию качества. Такая автоматизация также служит в качестве документации, отражающей необходимые проверки.

Примечание — Дополнительная информация о качестве данных для аналитики и машинного обучения приведена в *[21] — [24]*. Дополнительная информация о различных формах предвзятости и необъективности в данных, используемых в системах ИИ, приведена в *[25]*.

e) Документирование происхождения данных и истории их обработки и хранения

Поскольку обучающие данные могут определить поведение системы ИИ, важно знать их первоисточник, способ их обработки, их владельца и основания для их создания и сбора — на случай возникновения каких-либо проблем с данными или появления необходимости в их обновлении.

Метаданные о «родословной» данных (data lineage) документируют первоисточник данных, то, что с данными происходило впоследствии и как они перемещались во времени. Родословная данных позволяет видеть, как изменялись данные по мере прохождения процессов их обработки, одновременно сильно облегчая отслеживание первопричин ошибок в процессе анализа данных, а также отслеживание ошибок в версиях продукта в случае обнаружения проблем с исходными данными.

Метаданные о происхождении данных (data provenance) документируют факторы (лица и организации, системы и процессы), повлиявшие на данные, представляющие интерес, по существу сохраняя документированную историю данных и их первоначального происхождения.

Дополнительную информацию об управлении знаниями и об управлении информацией см. в 6.2.6 и 6.3.6 соответственно.

f) Очистка, объединение и подготовка данных

Подготовка данных — это набор операций над данными, которые приводят к желаемому результату. Он включает извлечение, объединение (слияние), очистку, фильтрацию, корректировку, дополнение, преобразование, кодирование и обработку отсутствующих значений.

Цель подготовки данных заключается в создании для данных признаков, которые используются в качестве входных данных для модели ИИ. Конструирование признаков представляет собой процесс выбора, описания и оптимизации признаков для их использования в модели ИИ. В рамках этого процесса выбор подходящих входных данных может быть сделан с использованием знаний о предметной области, путем анализа данных или экспериментирования с различными наборами признаков. Некоторые типы моделей включают оптимизацию выбора признаков. В целом чем меньше используется признаков, тем проще обучить модель МО, тем меньше рисков, связанных с ошибками в данных, и тем меньше усилий затрачивается на инженерию данных для ИИ.

Фильтрация удаляет нежелательные данные, которые:

- бесполезны для создания и/или верификации модели (это, например, выбросы в некоторых ситуациях);
- избыточны по объему, поэтому может быть достаточно выборки из них;
- вредны, поскольку вносят нежелательные предвзятость или дискриминацию (например, по признаку пола или этнической принадлежности);
- нарушают требования законодательства о защите персональных данных, предусматривающие удаление или деидентификацию (анонимизацию) персональных данных;
- представляют собой чувствительные данные, которые должны быть защищены от внутреннего или внешнего несанкционированного доступа.

В некоторых ситуациях аугментация (расширение) данных может помочь увеличить объем данных с целью создания более качественной модели или проведения большего количества тестов (например, путем поворота изображений).

Конвертирование (преобразование) данных и кодирование признаков используются для преобразования данных таким образом, чтобы удовлетворить критериям, которые модель ИИ устанавливает для входных данных (например, требование о том, что определенная переменная может принимать только значения «да» или «нет»).

Методы генеративного ИИ могут быть адаптированы для автоматического создания метаданных, поддерживающих модель ИИ, посредством выявления закономерностей в эксплуатационных данных.

Подготовка данных часто представляет собой поисково-исследовательский и, следовательно, менее структурированный процесс, включающий выполняемые вручную шаги и ситуативное создание кода. Подобный ситуативный (ad hoc) характер может затруднить повторяющееся проведение подготовки данных, и поэтому полезно стремиться к использованию многоразового автоматизированного процесса подготовки.

С учетом высокой сложности и поисково-исследовательского характера процесса подготовки данных важное значение имеет его (автоматизированное) тестирование.

g) Защита чувствительных данных

Некоторые аспекты систем ИИ полагаются на чувствительные данные, и когда это происходит, процесс инженерии данных для ИИ увеличивает поверхность атаки системы ИИ. Это означает, что помимо самой системы ИИ атаке могут быть подвергнуты элементы инженерии данных для ИИ — например, хранилище данных. Возникают риски безопасности и защиты персональных данных, особенно если данные о физических лицах собираются из различных источников. В подобных случаях необходимы осторожное обращение с данными и применение методов, обеспечивающих сохранение неприкосновенности частной жизни (защиту персональных данных).

6.4.9 Процесс реализации

6.4.9.1 Цель

Целью процесса реализации является реализация заданного элемента системы.

Данный процесс трансформирует требования, архитектуру и проектное решение, включая интерфейсы, в действия, которые создают элемент системы в соответствии с практикой выбранной технологии реализации для соответствующих технических специальностей или дисциплин. Результатом этого процесса является элемент системы, соответствующий заданным системным требованиям (включая выделенные и производные требования), архитектуре и проектному решению.

6.4.9.2 Результаты процесса

Применимы положения 6.4.7 *[1]* и 6.4.7 *[2]*, касающиеся результатов процесса.

В результате успешного выполнения части процесса реализации, связанной с инженерией модели ИИ:

a) создается работающая модель ИИ;

b) создается документация процесса создания модели.

6.4.9.3 Действия и задачи

Применимы положения 6.4.7 *[1]* и 6.4.7 *[2]*, касающиеся выполняемых действий и задач процесса, со следующим дополнением.

В рамках проекта необходимо реализовать следующие действия в соответствии с применимыми политиками и процедурами организации в отношении части процесса реализации, касающейся инженерии модели ИИ.

Для систем ИИ на основе машинного обучения добавляются следующие дополнительные действия:

a) Выбор алгоритма: выбор подходящего алгоритма машинного обучения с учетом типа задачи, выполняемой моделью (как, например, кластеризация, прогнозирование временных рядов, классификация), и лучше всего подходящего для решаемой задачи метода, который также может быть определен экспериментальным путем.

Одним из аспектов, которые следует принять во внимание при выборе (и настройке) алгоритмов, является вопрос о том, насколько интерпретируемой или объяснимой может быть модель. Как правило, наиболее эффективными оказываются те модели, которые сложнее интерпретировать. С другой стороны, интерпретируемые модели помогают укрепить доверие и обеспечить прозрачность. Такая прозрачность может быть полезна для обеспечения подотчетности, и она помогает разработчикам ИИ лучше понять предметную область и данные.

b) Обучение модели: алгоритм следует запускать на обучающих данных итеративно, чтобы сформировалось внутреннее представление (например, веса в нейронной сети). В случае обучения с учителем цель заключается в том, чтобы использовать примеры в составе обучающих данных для оценки

базовой функции, отображающей входные данные на желаемый результат (например, классифицирующей показанное на изображении животное как «кошку» или «собаку»). Важно, чтобы модель хорошо обобщала эти примеры, предотвращая перетренировку (overfitting), вследствие которой модель может хорошо работать на обучающих данных и плохо — на эксплуатационных данных.

c) Настройка модели: Применение методов оптимизации для поиска значений гиперпараметров, обеспечивающих наилучшую производительность, с использованием валидационных (проверочных) данных.

Для систем ИИ, основанных на инженерии знаний, добавляются следующие дополнительные действия:

d) Программирование знаний: после приобретения (см. 6.4.7) знания следует формализовать в рамках эвристической модели, в которой вычисления организуются либо явным образом (процедурный подход — более близкий к традиционному программированию), либо неявно, посредством установления правил и/или вероятностей (декларативный подход).

Следует определить и предписать комбинированную архитектуру, рассмотрев возможность использования облачных и периферийных вычислений для управления «возникающим» (эмерджентным) поведением систем ИИ, особенно в промышленных приложениях.

6.4.9.4 Особенности, специфические для ИИ

При реализации действий и задач данного процесса организациям следует учитывать следующие особенности, специфические для ИИ.

Системы ИИ можно рассматривать как традиционные программные системы, которые применяют одну или несколько моделей ИИ, поэтому в ходе их реализации используются те же практики, имеющие некоторые особенности, а также вводящие новые элементы. Примером может служить обычная хорошая практика работы с активно поддерживаемым списком согласованных работ и операций (перечнем невыполненных работ). Включение в такой перечень работ, связанных с ИИ, облегчает междисциплинарную координацию, планирование и оценку.

Модель ИИ обычно является частью приложения, которое, помимо самой модели, разрабатывается без какого-либо использования машинного обучения или инженерии знаний. Поскольку данные и знания играют в ИИ свои очень специфические роли, в настоящем документе инженерия модели ИИ включена как часть процесса реализации; а также включен отдельный процесс для инженерии данных для ИИ. Инженерия данных и инженерия моделей тесно взаимосвязаны, и многие действия по реализации сочетают в себе оба эти элемента, хотя они различны по своей природе.

В случае машинного обучения инженерия модели ИИ требует обучения модели с использованием обучающих данных. Это итеративная оптимизация, в ходе которой выбирается тип модели, настраиваются и изменяются ее гиперпараметры — до тех пор, пока модель не начнет адекватно работать на обучающем наборе данных. Таким образом, процесс инженерии данных для ИИ обычно связан с инженерией моделей ИИ и часто сильно зависит от опыта привлеченных экспертов. Автоматизированное машинное обучение представляет собой подход, при котором эти процессы автоматизируются полностью или частично с тем, чтобы уменьшить эту зависимость и сделать работу более эффективной. Эффективность также можно повысить, по возможности распределяя работу между экспертами и компьютерными ресурсами для проведения параллельного экспериментирования. В зависимости от используемых алгоритмов и применения автоматизированного МО для обучения моделей и осуществления иных оптимизаций могут потребоваться значительные вычислительные мощности и время.

Когда модель работает в соответствии со спецификациями прерываний (exceptions specifications) и/или заранее заданными спецификациями (predefined specifications), ее можно дополнительно настроить с использованием валидационных (проверочных) данных, а затем протестировать с использованием тестовых данных (см. 6.4.11).

Особым типом инженерии моделей является перенос обучения (трансферное обучение), при котором существующая модель МО применяется в качестве отправной точки для дальнейшего обучения для немного отличающегося варианта использования. Опираясь на предыдущие успехи в инженерии моделей, можно добиться повышения эффективности.

В процессе реализации можно опереться на существующие средства разработки программного обеспечения (software frameworks). Эти средства обычно предлагают различные встроенные модели ИИ и решения для обработки данных, для обучения моделей, тестирования и оркестровки.

В случае инженерии знаний инженерия моделей заключается в спецификации знаний в декларативной или процедурной форме. Знания приобретаются от экспертов (извлечение знаний) и/или посредством анализа данных (см. 6.4.8).

6.4.10 Процесс комплексирования

Применимы положения 6.4.8 *[1]* и 6.4.8 *[2]*, касающиеся цели, выполняемых действий, задач и результатов процесса.

6.4.11 Процесс верификации

6.4.11.1 Цель

Целью процесса верификации является обеспечение объективных доказательств того, что система или ее элементы удовлетворяют установленным для них требованиям и обладают заданными характеристиками.

В процессе верификации выявляются аномалии (ошибки, дефекты и сбои) во всех документах (таких, например, как системные требования или описание архитектуры), в реализованных элементах системы и в процессах жизненного цикла. Для этого используются соответствующие методы, технические способы, стандарты и правила. Данный процесс обеспечивает информацию, необходимую для определения способов устранения выявленных аномалий.

6.4.11.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.9 *[1]* и 6.4.9 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.11.3 Особенности, специфические для ИИ

Для процесса верификации дополнительные действия не определены. При выполнении действий и задач, указанных в 6.4.11.2, данный процесс должен быть расширен за рамки верификации систем с тем, чтобы принять во внимание упомянутые ниже особенности, характерные для ИИ.

Первая характерная для ИИ особенность, которую следует учитывать в процессе верификации, — это верификация через поведение.

В то время как традиционные системы программируются на поведение, в точности соответствующее предписанному, модели ИИ строятся таким образом, чтобы максимально приблизиться к желаемому поведению. Такой вероятностный характер означает, что верификацию этих моделей следует проводить с использованием статистических методов.

Поскольку модели ИИ трансформируют входные данные в выходные результаты, верификация моделей обычно осуществляется посредством использования наборов данных для верификации, содержащих входные данные и желаемые результаты, с применением статистических методов для оценки желаемой корректности и устойчивости к изменениям (см. описание процесса определения системных требований в 6.4.3). Наборы данных для верификации могут быть сформированы как из подмножества обучающих данных одного и того же источника, но при этом не использовавшихся для обучения модели, так и из данных, поступивших из иного источника. Преимущество последнего подхода заключается в том, что иной источник данных позволяет лучше протестировать способность модели к обобщению.

Можно выделить два типа наборов данных для верификации. Валидационные данные используются для выбора наилучшей модели среди моделей-кандидатов. Тестовые данные используются для установления того, адекватно ли функционирует и обобщает выбранная модель.

Второй характерной для ИИ особенностью, которую следует учесть в процессе верификации, является верификация посредством анализа.

Анализ кода является подходящим методом проверки, когда речь идет об исходном коде, специально написанном для системы ИИ, включая знания, которые представлены в коде в эвристических системах. Однако в случае моделей машинного обучения исходный код алгоритма не может быть проанализирован, если он является частью существующей библиотеки или платформы разработки. Поведение модели машинного обучения определяется ее представлением в виде набора параметров. Даже если такое представление модели и является читаемым, его правильность, как правило, очень трудно оценить, потому что работа алгоритма следует не шагам, реализованным программистами (следуя соответствующему человеческому мыслительному процессу), а шагам, которые были автоматически оптимизированы с целью максимизировать производительность модели.

Организации следует убедиться, что исходный код, имеющий отношение к ИИ, охватывается регулярными проверками таких аспектов качества кода, как сопровождаемость, тестируемость и возможность повторного использования (это могут быть, например, проводимый коллегами анализ сценариев обучения или же модульное тестирование кода для подготовки данных, аналогичные проверкам исходного кода, не имеющего отношения к ИИ).

Для получения более подробной информации о процессе непрерывной валидации см. 6.4.14.

6.4.12 Процесс переноса в среду промышленной эксплуатации

6.4.12.1 Цель

Целью процесса переноса в среду промышленной эксплуатации является обеспечение способности системы предоставлять услуги в среде эксплуатации в соответствии с требованиями заинтересованных сторон.

Данный процесс упорядоченным и планомерным образом переводит систему в состояние промышленной эксплуатации таким образом, чтобы система была функциональной, работоспособной и совместимой с другими системами, находящимися в промышленной эксплуатации. В рамках данного процесса в соответствии с соглашениями верифицированная система устанавливается вместе с релевантными вспомогательными системами (такими, например, как система планирования, система поддержки с персоналом технической поддержки, система обучения, система обучения пользователей). Процесс переноса в среду промышленной эксплуатации используется на каждом уровне в структуре системы и на каждой стадии для выполнения критериев, установленных для завершения стадии. Этот процесс включает в себя подготовку соответствующих систем обеспечения условий хранения, обработки и транспортировки.

6.4.12.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.10 *[1]* и 6.4.10 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.12.3 Особенности, специфические для ИИ

Для процесса переноса в среду промышленной эксплуатации дополнительные действия или задачи не определены. При выполнении действий и задач, указанных в 6.4.12.2, командам проектов и/или организациям следует принять во внимание упомянутые ниже особенности, характерные для ИИ.

Часто существует разница между реальной системой ИИ, в которой используется модель, и самой моделью, которая представляет собой конфигурацию или набор параметров (например, набор весовых коэффициентов нейронной сети).

Вследствие эксплуатационных требований модели ИИ могут быть развернуты в формате, отличном от того, в котором они были разработаны.

Организация должна стремиться к тому, чтобы поддерживать обновление моделей (проведение переобучения или использование инженерии знаний) и ведение постоянного мониторинга установленных метрик, связанных с использованием системы ИИ.

Организация должна оценить, каким образом может быть затронута производительность системы ИИ после того, как она будет введена в эксплуатацию, и, учитывая выявленные факторы, разработать соответствующие метрики для мониторинга. После развертывания система ИИ может демонстрировать неожиданное поведение (например, в результате предвзятости или из-за поступления на вход неожиданных данных). Ввиду этого мониторинг производительности важен, а процедуры и процессы могут быть расширенными по сравнению с традиционными системами.

На возможные сроки обновления модели влияют:

- изменения в соответствующих процессах функционирования;
- выявление изменений в соответствующих данных с течением времени (например, дрейф данных, как он описан в *[5]*);
- выявление снижения точности (например, вследствие дрейфа концепции, как объясняется в *[5]*);
- время, прошедшее с момента последнего обновления модели или с момента ее создания.

Поскольку некоторые системы ИИ способны со временем улучшать свою производительность, организация должна поддерживать достижение качества таких улучшений путем реализации процесса управления качеством. Например, отслеживание тенденций изменения объемов использования системы ИИ может помочь организации обеспечить ее непрерывное «качество при использовании».

6.4.13 Процесс валидации

6.4.13.1 Цель

Целью процесса валидации является обеспечение объективных доказательств того, что система в ходе ее использования выполняет поставленные перед ней деловые задачи и/или свою миссию, удовлетворяет требованиям заинтересованных сторон и реализует свое целевое назначение в целевой среде эксплуатации.

Целью валидации системы или ее элемента является обеспечение уверенности в ее способности выполнять свою миссию и/или реализовывать целевое назначение в определенных условиях эксплуатации. Итоги валидации утверждаются заинтересованными сторонами. Процесс валидации обеспечивает необходимую информацию для того, чтобы выявленные аномалии могли быть устранены соответствующим техническим процессом, в рамках которого аномалия возникла.

6.4.13.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.11 *[1]* и 6.4.11 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.13.3 Особенности, специфические для ИИ

Для процесса валидации дополнительные действия не определены. При выполнении действий и задач, указанных в 6.4.13.2, данный процесс должен быть расширен за рамки валидации систем с тем, чтобы принять во внимание упомянутые ниже особенности, характерные для ИИ.

Чтобы приобрести опыт работы с системами ИИ, организации могут выполнить проект проведения апробации (подтверждения работоспособности) концепции. В таком случае процесс валидации также включает валидацию самого ИИ, его полезности и рисков для организации в целом.

Организация должна стремиться к тому, чтобы поддерживать обновление моделей (проведение переобучения или использование инженерии знаний) и ведение постоянного мониторинга установленных метрик, связанных с использованием системы ИИ.

Организация должна оценить, каким образом может быть затронута производительность системы ИИ после того, как она будет введена в эксплуатацию, и, учитывая выявленные факторы, разработать соответствующие метрики для мониторинга. После развертывания система ИИ может демонстрировать неожиданное поведение (например, в результате предвзятости или из-за поступления на вход неожиданных данных). Ввиду этого важен мониторинг производительности, а процедуры и процессы могут быть расширены по сравнению с традиционными системами.

На возможные сроки обновления модели влияют:

- изменения в соответствующих процессах функционирования;
- выявление изменений в соответствующих данных с течением времени (например, дрейф данных, как он описан в *[5]*);
- выявление снижения точности (например, вследствие дрейфа концепции, как объясняется в *[5]*);
- время, прошедшее с момента последнего обновления модели или с момента ее создания.

Поскольку некоторые системы ИИ способны со временем улучшать свою производительность, организация должна поддерживать достижение качества таких улучшений путем реализации процесса управления качеством. Например, отслеживание тенденций изменения объемов использования системы ИИ может помочь организации обеспечить ее непрерывное «качество при использовании».

6.4.14 Процесс непрерывной валидации

6.4.14.1 Цель

Целью процесса непрерывной валидации является мониторинг того, чтобы модели ИИ с течением времени продолжали работать удовлетворительно и/или продолжали демонстрировать производительность модели ИИ.

Модели ИИ нацелены на моделирование желаемого поведения, которое может изменяться со временем. Кроме того, со временем могут меняться эксплуатационные данные. По этой причине важно измерять и вести мониторинг отклонений входных данных (дрейф данных) и отклонений, влияющих на целевой результат (дрейф концепции) с использованием тестовых данных. Данный процесс является расширением процесса обеспечения уверенности в качестве (см. 6.3.8).

Если отклонения существенны, то в случае машинного обучения требуется провести переобучение и/или организовать непрерывное обучение в рамках процесса сопровождения (технической поддержки) (см. 6.4.16). Наличие отклонений также может указывать на другие проблемы, например на проблемы с качеством данных или на сбои в работе системы. Если система ИИ применяет автоматическое непрерывное обучение без участия человека, то в него следует включить автоматический процесс отката при достижении определенных пороговых значений с тем, чтобы предотвратить нежелательные изменения модели.

6.4.14.2 Результаты процесса

В результате успешного выполнения процесса непрерывной валидации:

a) результаты валидации должны быть задокументированы в журнале валидации;

b) может быть принято решение о проведении технического обслуживания модели ИИ (ее переобучения).

6.4.14.3 Действия и задачи

В рамках проекта в соответствии с применимыми политиками и процедурами организации в отношении процесса непрерывной валидации необходимо реализовать следующие действия:

a) провести мониторинг дрейфа данных посредством выполнения проверок входных данных модели с тем, чтобы определить, не отклоняются ли они от тех, на которых модель была обучена;

b) провести мониторинг дрейфа концепции посредством измерения производительности модели с использованием обновленных тестовых данных, или посредством выявления каких-либо аномалий в выходных значениях или в распределении выходных значений — например, сравнивая недавние выходные данные с теми, что были получены ранее;

c) провести мониторинг других показателей и характеристик, изменения которых со временем можно ожидать (см. 6.4.3), например время выполнения, прозрачность и справедливость;

d) в случае отклонений принять решение о том, следует ли проводить техническое обслуживание модели ИИ;

e) в случае отклонений применить «защитные ограждения», если таковые были определены путем установления границ для выходных данных, или же перейти на использование по умолчанию альтернативной безопасной модели;

f) определить частоту проведения валидации.

6.4.15 Процесс функционирования

6.4.15.1 Цель

Целью процесса функционирования является использование системы для предоставления ею своих сервисов.

В рамках данного процесса устанавливаются требования и назначается персонал для работы с системой, проводится мониторинг сервисов и оценивается эффективность работы операторов с системой. Для поддержания соответствующих сервисов выявляются и анализируются аномалии функционирования в сравнении с соглашениями, требованиями заинтересованных сторон и организационными ограничениями.

6.4.15.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.12 *[1]* и 6.4.12 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.15.3 Особенности, специфические для ИИ

Для процесса функционирования дополнительные действия не определены. При выполнении действий и задач, указанных в 6.4.15.2, данный процесс должен быть расширен за рамки функционирования систем для того, чтобы принять во внимание следующие особенности, специфические для ИИ.

Первая такая особенность, принимаемая во внимание в процессе функционирования, связана с использованием вычислительных ресурсов и энергии.

Системы ИИ могут потреблять значительные вычислительные мощности и объемы памяти, особенно при обучении моделей машинного обучения (в зависимости от типа алгоритма). Иногда для ускорения обработки используется специализированное оборудование — например, графические процессоры благодаря наличию у них возможностей для массовой параллельной обработки. Возникающие в результате дополнительные затраты и «углеродный след» могут стать существенными факторами при принятии решений, касающихся частоты проведения обучения, выбора алгоритма или же использования машинного обучения в целом.

Модели развертываются для работы либо в пакетном, либо в непрерывном режимах, в зависимости от того, есть ли у системы ИИ постоянная потребность в результатах работы модели. Работающие в непрерывном режиме модели обычно имеют более строгие требования к эффективности функционирования.

Вторая характерная для ИИ особенность заключается в том, что организации уже на ранних стадиях жизненного цикла следует принимать во внимание эксплуатационные данные, с которыми будет работать система ИИ. В число рассматриваемых вопросов могут входить доступность, пригодность для обеспечения желаемого поведения, многообразие признаков, согласованность между обучающими, тестовыми и эксплуатационными данными при одновременном использовании, когда это необходимо, независимых наборов данных. Такие соображения могут, например, быть трансформированы в функциональные и технические спецификации, руководства пользователя и/или пользовательские спецификации, или же в метрики для эксплуатационных данных.

Третьей специфической для ИИ особенностью, принимаемой во внимание в процессе функционирования, является развертывание модели данной системы ИИ.

Модели могут быть развернуты отдельно в зависимости от конкретных требований к среде исполнения в аппаратной и/или программной части. Это необходимо, когда модели заменяются чаще, чем остальная часть системы — например, после проведения повторного обучения.

В некоторых ситуациях исполняемые модели отличаются от моделей, которые используются в ходе разработки по причине того, что среда разработки может не соответствовать требованиям среды

исполнения. Примером может служить ситуация, когда модель развертывается во встроенной системе с ограниченным набором поддерживаемых технологий, однако разрабатывается в эмулируемой среде на персональном компьютере.

Организация должна стремиться к тому, чтобы поддерживать обновление моделей (путем переобучения или применения инженерии знаний) и проводить непрерывный мониторинг установленных метрик, связанных с использованием системы ИИ.

Организация должна оценить, каким образом может быть затронута производительность системы ИИ после того, как она начнет функционировать, и, учитывая выявленные факторы, разработать соответствующие метрики для мониторинга. После развертывания система ИИ может демонстрировать неожиданное поведение (например, в результате предвзятости или из-за поступления на вход неожиданных данных). Поэтому мониторинг производительности важен, а процедуры и процессы могут быть более обширными по сравнению с традиционными системами.

На возможные сроки обновления модели влияют:

- изменения в соответствующих процессах функционирования;
- выявление изменений в соответствующих данных с течением времени (например, дрейф данных, как он описан в *[5]*);
- выявление снижения точности (например, вследствие дрейфа концепции, как объясняется в *[5]*);
- время, прошедшее с момента последнего обновления модели или с момента ее создания.

Поскольку некоторые системы ИИ способны со временем улучшать свою производительность, организация должна добиваться качества таких улучшений путем реализации процесса управления качеством. Например, отслеживание тенденций изменения объемов использования системы ИИ может помочь организации обеспечить ее постоянное «качество при использовании». Кроме того, следует сообщать обо всех инцидентах, включая системные сбои и ошибки в данных, и давать им оценку.

6.4.16 Процесс сопровождения

6.4.16.1 Цель

Целью процесса сопровождения является поддержание способности системы выполнять заданные функции.

В рамках данного процесса контролируется способность системы выполнять заданные функции, документируются инциденты с целью их анализа, предпринимаются действия по корректировке, адаптации, исправлению и предупреждению нарушений функционирования, а также подтверждается восстановленная работоспособность.

Примечание — Подробные сведения о типах действий (по корректировке, адаптации, исправлению и предупреждению нарушений функционирования), выполняемых в рамках процесса сопровождения, можно найти в *[26]*.

6.4.16.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.13 *[1]* и 6.4.13 *[2]*, касающиеся выполняемых действий и задач, а также результатов процесса, со следующим дополнением. В рамках проекта необходимо реализовать следующие действия в соответствии с применимыми политиками и процедурами организации в отношении процесса сопровождения:

- непрерывное обучение модели ИИ: в отличие от повторного обучения модели, проводимого время от времени по мере необходимости, непрерывное обучение означает, что модель постоянно эволюционирует по мере того, как она обучается на эксплуатационных данных. Цель непрерывного обучения заключается в том, чтобы приспособиться к изменению желаемого поведения с течением времени, связанного с изменениями во входных данных и соответствующими изменениями в выходных результатах. Непрерывное обучение может быть реализовано как в форме регулярно проводимого автоматизированного переобучения, так и в форме инкрементального обучения, если алгоритм обучения его поддерживает;
- организация оценивает, каким образом может быть затронута производительность системы ИИ после начала использования, и, учитывая выявленные факторы, разрабатывает соответствующие метрики для мониторинга, которые будут использоваться в процессе непрерывной валидации (см. 6.4.14);
- организация стремится поддерживать обновление модели (путем переобучения или применения инженерии знаний) для поддержания ее производительности, которая подлежит мониторингу в рамках процесса непрерывной валидации (см. 6.4.14).

6.4.16.3 Особенности, специфические для ИИ

Процесс сопровождения, описанный в 6.4.13 *[1]* и 6.4.13 *[2]* должен быть расширен для того, чтобы охватить действия по непрерывному обучению.

Как и в случае разработки традиционных программных систем, процесс сопровождения может охватывать все действия, которые выполняются в более ранних процессах, особенно в процессе реализации. Проектное решение и его реализация могут постоянно эволюционировать. То же самое верно и для систем ИИ. Могут потребоваться переобучение или иные обновления моделей (об инженерии моделей как части процесса реализации см. 6.4.9). Могут быть собраны новые обучающие данные и изменен процесс подготовки данных с тем, чтобы система продолжала должным образом функционировать при изменении внешнего окружения и/или требований (см. 6.4.8). Также может потребоваться обновление тестовых данных (см. 6.4.11). Вместо проводимого время от времени переобучения может идти речь о непрерывном обучении.

Частью процесса сопровождения является мониторинг системы. Поскольку мониторинг модели ИИ сильно отличается от типичного мониторинга системы, выделен отдельный процесс непрерывной валидации (см. 6.4.14).

После развертывания система ИИ может демонстрировать неожиданное поведение (например, в результате предвзятости или из-за поступления на вход неожиданных данных), поэтому мониторинг производительности важен, а процедуры и процессы могут быть более обширными по сравнению с традиционными системами.

На возможные сроки обновления модели влияют:

- изменения в соответствующих процессах функционирования;
- выявление изменений в соответствующих данных с течением времени (например, дрейф данных, как он описан в *[5]*);
- выявление снижения точности (например, вследствие дрейфа концепции, как объясняется в *[5]*);
- время, прошедшее с момента последнего обновления модели или с момента ее создания.

Поскольку некоторые системы ИИ способны со временем улучшать свою производительность, организация должна поддерживать такие улучшения путем реализации процесса управления качеством. Например, отслеживание тенденций изменения объемов использования системы ИИ может помочь организации обеспечить ее неснижаемое «качество при использовании».

Если вследствие изменений в эксплуатационных данных или выявленной предвзятости развернутая модель работает неоптимально, ее можно откатить до более ранней версии, работавшей лучше, внести исправления и/или сделать ее более надежной и устойчивой к изменениям. Таким образом, автоматизированный процесс отката модели является полезным для быстрого решения проблемы неоптимальной производительности модели.

Следует учитывать, что сопровождение системы ИИ может быть сложным, поскольку поведение системы ИИ может быть нестабильным и не всегда объяснимым — даже при наличии инженерной документации. Кроме того, версии элементов конфигурации не всегда отражают поведение системы ИИ.

6.4.17 Процесс изъятия и списания

6.4.17.1 Цель

Целями процесса изъятия и списания являются: прекращение функционирования системы или ее элементов для предопределенных целей; надлежащее обращение с замененными или выведенными из эксплуатации элементами и должное внимание к выявленным критически важным потребностям, связанным с выводом из эксплуатации или уничтожением (например, в соответствии с соглашением, политикой организации или в связи с экологическими и правовыми вопросами, а также вопросами обеспечения безопасности и защищенности).

6.4.17.2 Выполняемые действия, задачи и результаты процесса

Применимы положения 6.4.14 *[1]* и 6.4.14 *[2]*, касающиеся выполняемых действий, задач и результатов процесса.

6.4.17.3 Особенности, специфические для ИИ

Процесс, описанный в 6.4.14 *[1]* и 6.4.14 *[2]*, должен быть расширен за рамки изъятия и списания систем и охватывать уничтожение данных либо их передачу другой организации. Этот новый вид деятельности может привести к новым проблемам, поскольку для любых данных, связанных с системой, может потребоваться проведение их тщательного изъятия и списания ввиду рисков для безопасности или неприкосновенности частной жизни. Кроме того, процесс изъятия и списания данных должен принимать во внимание соответствующие требования к срокам их хранения, как это предусмотрено стандартом *[16]*.

Приложение А
(справочное)

Наблюдения, основанные на анализе вариантов использования

(см. [13])

А.1 Особенности специфических для систем ИИ процессов жизненного цикла по сравнению с традиционными системами

А.1.1 Общие положения

Существуют международные стандарты для жизненного цикла, в частности *[1]* и *[2]*. По итогам анализа вариантов использования можно сделать заключение о том, что среди тридцати процессов, определенных в *[1]* и *[2]*, у семи упомянутых ниже процессов наблюдаются специфические для ИИ особенности. Для каждого из этих семи процессов приведены примеры из *[13]*.

А.1.2 Процесс управления информацией

Согласно *[1]* целью процесса управления информацией является «создание, сбор, подтверждение, преобразование, хранение, извлечение, распространение, списание и изъятие информации для соответствующих заинтересованных сторон».

Примеры использования включают:

- выявление выбросов, выбор признаков и заполнение отсутствующих значений выполняются на стадии предварительной обработки в варианте использования 24 (решение ИИ для прогнозирования послеоперационной остроты зрения при выполнении операций лазерной коррекции зрения по методике LASIK);
- аугментация данных выполняется на стадии предварительной обработки в варианте использования 42 (сервис ИИ обслуживания клиентов с учетом их эмоций);
- создание обучающих выборок (путем разметки данных) и их предварительная обработка (путем сегментации предложений и создания векторов слов) выполняются на стадии обучения в варианте использования 43 (распознавание намерений пользователя на основе глубокого обучения).

А.1.3 Процесс реализации

Согласно *[1]* целью процесса реализации является «создание заданного элемента системы».

Примеры вариантов использования включают:

- обучение моделей, которое выполняется при глубоком обучении на соответствующей стадии в следующих вариантах использования:
  - объяснимый искусственный интеллект для геномной медицины (вариант использования 1);
  - решение ИИ для быстрого выявления дефектов в процессе контроля качества лопастей ветряных турбин (вариант использования 4);
  - извлечение информации из размеченных вручную производственных контрольных листов (вариант использования 21);
  - повышение эффективности управления дорожным движением и точности выявления нарушений с помощью технологий ИИ (вариант использования 29);
  - сервис обслуживания клиентов с помощью технологий ИИ с учетом их эмоций (вариант использования 42);
  - распознавание намерений пользователя на основе глубокого обучения (вариант использования 43);
  - решение ИИ для оптимизации управления сигналами светофоров на основе объединения данных из нескольких источников (вариант использования 49);
  - решение ИИ для контроля качества электронных медицинских документов в режиме реального времени (вариант использования 50);
- регрессия с помощью деревьев принятия решений с градиентным усилением, которая выполняется на стадии обучения в варианте использования 24 (решение ИИ для прогнозирования послеоперационной остроты зрения при выполнении операций лазерной коррекции зрения по методике LASIK).

А.1.4 Процесс верификации

Согласно *[1]* целью процесса верификации является «объективное подтверждение того, что система или ее элементы удовлетворяют установленным требованиям и обладают заданными характеристиками».

В качестве вариантов использования можно привести следующие:

- проводится оценка ключевых показателей эффективности в ходе слепого тестирования на стадии оценки в варианте использования 4 (решение ИИ для быстрого выявления дефектов в процессе контроля качества лопастей ветряных турбин). Если удовлетворяется определенное условие (например, покрытие 1 составляет 95 % или более, а разделение 2 составляет 20 % или менее), то следует переход к стадии выполнения;

- производительность (представленная такими ключевыми показателями эффективности, как точность и отклик модели) анализируется на стадии оценки в варианте использования 42 (сервис обслуживания клиентов с помощью ИИ с учетом их эмоций). Если показатели производительности сравнимы с результатами, демонстрируемыми современными методами на открытых наборах данных, и соответствуют заданным условиям на собственных наборах данных, то следует переход к стадии выполнения.

А.1.5 Процесс переноса в среду промышленной эксплуатации

Согласно *[1]* целью процесса переноса в среду промышленной эксплуатации является «обеспечение способности системы предоставлять сервисы в среде эксплуатации в соответствии с требованиями заинтересованных сторон».

В техническом отчете *[13]* нет вариантов использования, относящихся к процессу переноса в среду промышленной эксплуатации.

Во многих вариантах использования итоговая модель требует развертывания в среде эксплуатации технических средств исполнения в иной конфигурации в сравнении с той, что применялась в процессе реализации.

А.1.6 Процесс валидации

Согласно *[1]* целью процесса валидации является «объективное подтверждение того, что выполнение миссии или решение поставленных задач, обеспечиваемые системой при ее использовании в предполагаемой среде эксплуатации, соответствует требованиям заинтересованных сторон».

Среди вариантов использования в техническом отчете *[13]* подходящих примеров выявлено не было.

А.1.7 Процесс функционирования

Согласно *[1]* целью процесса функционирования является «использование системы для предоставления ею своих сервисов».

Примером может служить вариант использования 1 (объяснимый искусственный интеллект для геномной медицины), когда помимо прогноза также представляется и его объяснение.

А.1.8 Процесс сопровождения

Согласно ИСО/МЭК/ИИЭР 12207 *[1]* целью процесса сопровождения является «поддержание способности системы выполнять заданные функции».

Примеры вариантов использования при переобучении в процессе сопровождения включают:

- решение ИИ для быстрого выявления дефектов в процессе контроля качества лопастей ветряных турбин (вариант использования 4);
- решение ИИ для прогнозирования послеоперационной остроты зрения при выполнении операций лазерной коррекции зрения по методике LASIK (вариант использования 24);
- сервис обслуживания клиентов с использованием ИИ с учетом их эмоций (вариант использования 42);
- распознавание намерений пользователя на основе глубокого обучения (вариант использования 43).

А.2 Поток процессов, специфических для ИИ

В вариантах использования наблюдался описанный ниже поток процессов. Название процесса соответствует стадии в шаблоне описания варианта использования, а в круглых скобках приведено название процесса в соответствии с *[1]*. Рисунок А.1 показывает поток процессов, специфических для ИИ.

Предварительная обработка (процесс управления информацией) → Обучение (процесс реализации) → Оценка (процесс верификации → процесс переноса в среду промышленной эксплуатации → процесс валидации) → Выполнение (процесс функционирования) → Предварительная обработка (процесс управления информацией) → Переобучение (процесс сопровождения) → Оценка (процесс верификации) → ...

Процессы Обучение (процесс реализации) и Переобучение (процесс сопровождения) можно повторять до тех пор, пока не будет успешно пройдена Оценка (процесс верификации). Процесс Выполнение (процесс функционирования) повторяется с различными входными данными, и другой процесс (такой, как процесс объяснения) может выполняться в дополнение к основному процессу, ключевыми показателями эффективности (KPI) которого являются:

- охват (coverage): соотношение дефектов, включенных или обнаруженных в областях продукта, которые «представляют интерес» для ручной проверки;
- разделение (split): прогнозируемая доля областей продукта, которые «представляют интерес» для ручной проверки.

Этот поток процессов можно проиллюстрировать с помощью рисунка А.1, который ссылается на диаграмму «Поток на стадиях обучения и использования» в *[27]*.

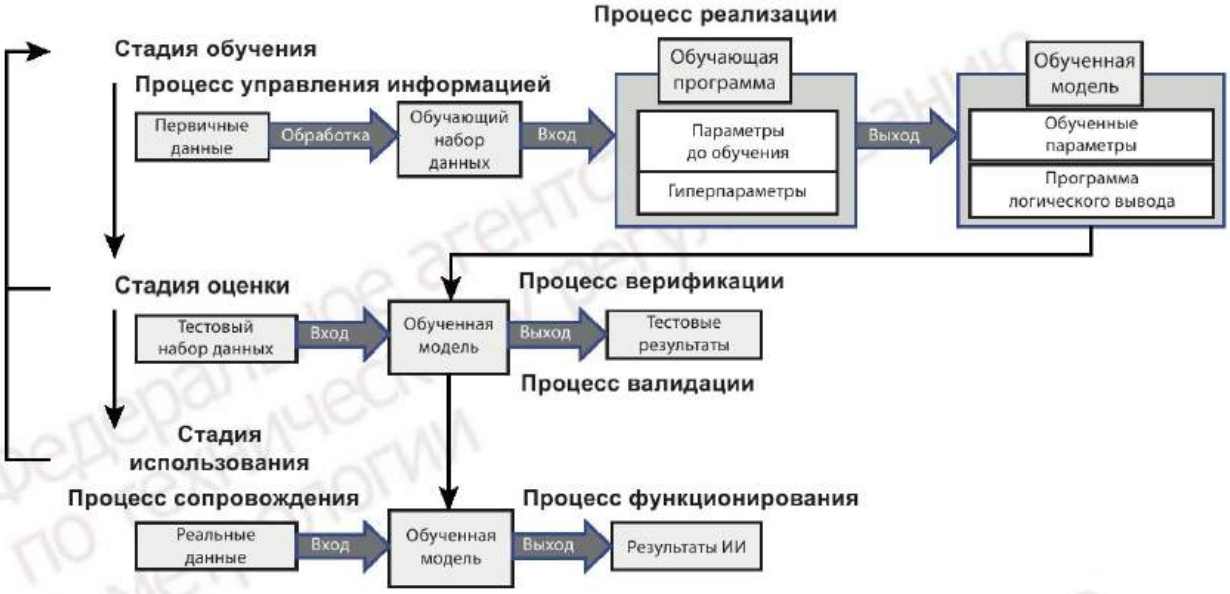

Рисунок А.1 — Поток процессов, специфических для ИИ

А.3 Данные для управления потоком процессов

Считается, что данные, описанные в «Постусловии» раздела «Сценарий процесса» шаблона описания варианта использования, управляют потоком процессов. Ниже приведены примеры из собранных вариантов использования:

- соблюдение требования к точности прогноза (например, точность прогноза должна составлять не менее 90 %) является критерием «успеха», позволяющего перейти от процесса Оценка (процесс верификации) к процессу Выполнение (процесс функционирования) в следующих вариантах:
  - объяснимый искусственный интеллект для геномной медицины (вариант использования 1),
  - извлечение информации из размеченных вручную производственных контрольных листов (вариант использования 21),
  - распознавание намерений пользователя на основе глубокого обучения (вариант использования 43);
- преимущество новой модели над старой при их сравнении с использованием АВ-теста является критерием «успеха», позволяющим перейти от процесса Переобучение (процесс сопровождения (технической поддержки)) к процессу Оценка (процесс верификации) в варианте использования 43 (распознавание намерений пользователя на основе глубокого обучения);
- комбинация двух ключевых показателей эффективности — охвата и разделения — используется для управления потоком от процесса Оценка (процесс верификации) до процесса Выполнение (процесс функционирования) в варианте использования 4 (решение ИИ для быстрого выявления дефектов в процессе контроля качества лопастей ветряных турбин).

Приложение ДА
(справочное)

Сведения о соответствии ссылочных национальных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте

Таблица ДА.1

Обозначение ссылочного национального стандарта	Степень соответствия	Обозначение и наименование соответствующего международного стандарта
ГОСТ Р 71476—2024 (ИСО/МЭК 22989:2022)	MOD	ISO/IEC 22989:2022 «Информационные технологии. Искусственный интеллект. Концепции и терминология искусственного интеллекта»

Примечание — В настоящей таблице использовано следующее условное обозначение степени соответствия стандарта:
- MOD — модифицированный стандарт.

Библиография

[1]	ISO/IEC/IEEE 12207:2017	Системная и программная инженерия. Процессы жизненного цикла программных средств (Systems and software engineering — Software life cycle processes)
[2]	ISO/IEC/IEEE 15288:2023	Системная и программная инженерия. Процессы жизненного цикла системы (Systems and software engineering — System life cycle processes)
[3]	ISO/IEC TR 5469:2024	Искусственный интеллект. Функциональная безопасность и ИИ-системах (Artificial intelligence — Functional safety and AI systems)
[4]	ИСО/МЭК 42001:2023	Информационные технологии. Искусственный интеллект. Система менеджмента (Information Technology — Artificial intelligence — Management system)
[5]	ИСО/МЭК 23053:2022	Платформа разработки систем искусственного интеллекта (ИИ) с использованием машинного обучения (МО) [Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)]
[6]	ISO/IEC TR 24368:2022	Информационные технологии. Искусственный интеллект. Обзор этических и социальных проблем (Information technology — Artificial intelligence — Overview of ethical and societal concerns)
[7]	ИСО/МЭК 23894:2023	Информационные технологии. Искусственный интеллект. Руководство по менеджменту риска (Information technology — Artificial intelligence — Guidance on risk management)
[8]	IEEE 7000—2021	Стандартный типовой процесс IEEE для решения этических проблем при проектировании систем (IEEE Standard Model Process for Addressing Ethical Concerns during System Design)
[9]	ИСО/МЭК 5339:2024	Информационные технологии. Искусственный интеллект. Руководство по ИИ-приложениям (Information technology — Artificial intelligence — Guidance for AI applications)
[10]	ISO/IEC/IEEE 15289:2019	Системная и программная инженерия. Содержание информационных продуктов процесса жизненного цикла (документация) [Systems and software engineering — Content of life-cycle information items (documentation)]
[11]	МЭК 62304:2006 + Amd.1:2015	Программное обеспечение медицинских изделий. Процессы жизненного цикла (Medical device software — Software life cycle processes)
[12]	ISO/IEC/IEEE 24748-1:2018	Системная и программная инженерия. Управление жизненным циклом. Часть 1. Руководство по управлению жизненным циклом (Systems and software engineering — Life cycle management — Part 1: Guidelines for life cycle management)
[13]	ISO/IEC TR 24030:2021	Информационные технологии. Искусственный интеллект (ИИ). Примеры использования [Information technology — Artificial intelligence (AI) — Use cases]
[14]	ИСО/МЭК 25059:2023	Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модель качества для систем на основе ИИ [Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems]
[15]	ISO/IEC TS 25058:2024	Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Руководство по оценке качества ИИ-систем [Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of artificial intelligence (AI) systems]
[16]	ИСО/МЭК 38507:2022	Информационные технологии. Стратегическое управление ИТ. Последствия использования искусственного интеллекта организациями для стратегического управления (Information Technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations)
[17]	ИСО 14971:2019	Изделия медицинские. Применение менеджмента риска к медицинским изделиям (Medical devices — Application of risk management to medical devices)

[18] ИСО 10007:2017 Менеджмент качества. Руководство по менеджменту конфигурации (Quality management — Guidelines for configuration management)

[19] Dwork Cynthia, McSherry Frank, Nissim Kobbi, Smith Adam, Calibrating Noise to Sensitivity in Private Data Analysis. Berlin: Springer, 2006

[20] ИСО/МЭК 5392:2024 Информационные технологии. Искусственный интеллект. Эталонная архитектура инженерии знаний (Information Technology — Artificial Intelligence — Reference architecture of knowledge engineering)

[21] ИСО/МЭК 5259-1:2024 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 1. Обзор, терминология и примеры [Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples]

[22] ISO/IEC FDIS 5259-2 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 2. Показатели качества данных [Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures]

[23] ИСО/МЭК 5259-3:2024 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 3. Требования и рекомендации по управлению качеством данных [Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines]

[24] ИСО/МЭК 5259-4:2024 Искусственный интеллект. Качество данных для аналитики и машинного обучения. Часть 4. Рамочная концепция процесса обеспечения качества данных [Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework]

[25] ISO/IEC TR 24027:2021 Информационные технологии. Искусственный интеллект (ИИ). Предвзятость в ИИ-системах и принятие решений с помощью ИИ [Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making]

[26] ISO/IEC/IEEE 14764:2022 Разработка программного обеспечения. Процессы жизненного цикла программного обеспечения. Сопровождение (Software engineering — Software life cycle processes — Maintenance)

[27] Ministry of Economy, Trade and Industry of Japan Contract Guidelines on Utilization of AI and Data, 9 December 2019

[28] Serban A., van der Blom K., Hoos H., Visser J. Engineering best practices for Machine Learning, October 2020 (3) 1—2

УДК 004.8:006.352 ОКС 35.020

Ключевые слова: искусственный интеллект, жизненный цикл систем искусственного интеллекта, система искусственного интеллекта

Технический редактор *И.Е. Черепкова*
Корректор *Л.С. Лысенко*
Компьютерная верстка *Л.А. Круговой*

Сдано в набор 31.10.2024. Подписано в печать 20.11.2024. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 4,65. Уч.-изд. л. 4,16.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

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