ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р ИСО/МЭК 22989–2022

Информационные технологии

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Концепции и терминология искусственного интеллекта

(ISO/IEC 22989:2022, MOD)

Издание официальное

Настоящий проект стандарта не подлежит применению до его утверждения

Москва
Российский институт стандартизации
2023

Предисловие

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

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

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от__ _____ 202_ г. № ____

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 22989:2022 «Информационные технологии – Искусственный интеллект – Концепции и терминология искусственного интеллекта» (ISO/IEC 22989:2022 Information technology – Artificial intelligence – Artificial intelligence concepts and terminology)

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.gost.ru)

Содержание

Введение

Рост вычислительных мощностей, снижение стоимости вычислений, доступность больших объёмов данных из многочисленных источников, недорогие курсы онлайн-обучения и алгоритмы, способные достигать или превосходить по скорости и точности уровень производительности человека при выполнении конкретных задач, сделали возможным практическое применение искусственного интеллекта (ИИ), делая ИИ всё более важным направлением информационных технологий.

Искусственный интеллект - это междисциплинарная область, широко опирающаяся на информатику, науку о данных, естественные и гуманитарные науки, математику, общественные науки и другие дисциплины. В настоящем документе широко используются такие термины, как «интеллектуальный», «интеллект», «понимание», «знания», «обучение», «решения», «навыки» и т.д., однако целью документа является не «очеловечивание» ИИ-систем, а отражение того факта, что некоторые ИИ-системы могут рудиментарно имитировать подобные характеристики.

Существует множество предметных областей ИИ-технологии. Эти предметные области тесно взаимосвязаны между собой и быстро развиваются, поэтому сложно отразить актуальность всех таких технических областей на единой карте. Исследования ИИ охватывают такие аспекты, как «обучение, распознавание и предсказание», «вывод, знание и язык» и «выявление, поиск и создание». В этих исследованиях также рассматриваются взаимозависимости между данными аспектами [23].

Представление об ИИ как о потоке процессов ввода и вывода разделяется многими исследователями ИИ, и исследования каждого этапа этого процесса продолжаются. Стандартизированные концепции и

терминология необходимы заинтересованным в данной технологии сторонам для лучшего понимания и принятия технологии более широкой аудиторией. Кроме того, концепции и категории ИИ дают возможность сравнивать и классифицировать различные решения по таким свойствам, как доверие, робастность, жизнеспособность, надёжность, точность, безопасность, защищенность и обеспечение защиты персональных данных. Это позволяет заинтересованным сторонам выбирать подходящие решения для своих приложений и сравнивать качество доступных на рынке решений.

Поскольку в настоящем стандарте термин ИИ определяется только в смысле дисциплины, то контекст его использования можно описать следующим образом: ИИ – область науки и техники, рассматривающая технические системы, которые порождают такие результаты, как контент, прогнозы, рекомендации или решения для заданного набора поставленных человеком задач.

В данном стандарте содержатся стандартизированные концепции и терминология, которые должны помочь более широкому кругу заинтересованных сторон лучше понять и использовать ИИ-технологию. Стандарт предназначен для широкой аудитории, включающей как экспертов, так и лиц, не имеющих соответствующего практического опыта, - в то же время читать некоторые конкретные разделы может быть проще при наличии более основательных знаний в области информатики. Это в первую очередь касается разделов 5.10, 5.11 и 8, которые носят более технический характер, чем остальная часть стандарта.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Информационные технологии

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Концепции и терминология искусственного интеллекта

Information technology – Artificial intelligence – Artificial intelligence concepts and terminology

Дата введения – 2023 -__-__

1 Область применения

Настоящий стандарт определяет терминологию и описывает концепции в области искусственного интеллекта (ИИ).

Данный стандарт можно использовать при разработке других стандартов и для поддержки обмена информацией между различными заинтересованными сторонами.

Данный стандарт применим в организациях любого типа (например, в коммерческим организациях, в государственных учреждениях, в некоммерческих организациях).

2 Нормативные ссылки

В настоящем стандарте нормативные ссылки отсутствуют.

Издание официальное

3 Термины и определения

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

- платформа ИСО для онлайн-просмотра материалов по стандартам (Online Browsing Platform, OBP) доступна по адресу https://www.iso.org/obp/ui

- база данных МЭК «Электропедия» (IEC Electropedia) доступна по адресу http://www.electropedia.org/

В настоящем стандарте применены следующие термины с соответствующими определениями.

3.1 Термины, относящиеся к искусственному интеллекту

3.1.1. **ИИ-агент** (AI agent): автоматически (3.1.7) действующий объект, который воспринимает своё окружение, реагирует на него, а также предпринимает действия для достижения своих целей.

3.1.2. **ИИ-компонент** (AI component): один из функциональных элементов, из которых построена ИИ-система (3.1.4).

3.1.3. **искусственный интеллект, ИИ** (artificial intelligence, AI): <дисциплина> исследование и разработка механизмов и приложений ИИ-систем (3.1.4)

Примечание – Исследования и разработки могут проводиться в одной или нескольких областях, таких как информатика, наука о данных, гуманитарные науки, математика и естественные науки

3.1.4. **система искусственного интеллекта; ИИ-система** (artificial intelligence system, AI system): техническая система, которая порождает такие конечные результаты, как контент, прогнозы, рекомендации или решения для заданного набора определенных человеком целей.

Примечание 1 – В технической системе могут применяться различные связанные с искусственным интеллектом (3.1.3) методы и подходы для разработки модели (3.1.23) для представления данных, знаний (3.1.21), процессов и т. д., которая может быть использована для решения задач (3.1.35).

Примечание 2 – ИИ-системы проектируются для эксплуатации с различными уровнями автоматизации (3.1.7).

3.1.5. **автономность; автономный** (autonomy, autonomous): способность системы изменять свою целевую область использования и/или цель без внешнего вмешательства, контроля или надзора.

3.1.6. **специализированная интегральная схема; интегральная схема специального назначения** (application specific integrated circuit, ASIC): интегральная схема, специализированная под конкретное применение.

[ИСО/МЭК/ИИЭР 24765:2017, 3.193]

3.1.7. **автоматический, автоматизированный; автоматизация** (соответственно automatic, automated, automation) характеристика процесса или системы, которые при определённых условиях функционируют без вмешательства человека.

[ИСО/МЭК 2382:2015, 2121282]

3.1.8. **когнитивные вычисления** (cognitive computing): категория ИИ-систем (3.1.4), обеспечивающих более естественное взаимодействие людей с машинами.

Примечание – Решаемые при помощи когнитивных вычислений задачи связаны с машинным обучением (3.3.5), обработкой речи, обработкой естественного языка (3.6.9), компьютерным зрением (3.7.1) и человеко-машинными интерфейсами.

3.1.9. **непрерывное обучение; продолжающееся обучение; инкрементальное обучение на стадии эксплуатации** (continuous learning, continual learning, lifelong learning): инкрементальное (пошаговое, последовательное) обучение ИИ-системы (3.1.4), которое продолжается на постоянной основе в течение всей стадии эксплуатации в жизненном цикле ИИ-системы.

3.1.10. **коннекционизм, парадигма коннекционизма, коннекционистская модель, коннекционистский подход** (connectionism, connectionist paradigm, connectionist model, connectionist approach): форма когнитивного моделирования, при котором используется сеть взаимосвязанных простых вычислительных элементов.

3.1.11. **интеллектуальный анализ данных, извлечение знаний из данных** (data mining): вычислительный процесс, который выявляет закономерности и тенденции посредством анализа количественных данных в разных разрезах и с различных точек зрения; проводит их категоризацию и сводит воедино потенциальные взаимосвязи и воздействия.

[ИСО 16439:2014, 3.13]

3.1.12. **декларативные знания** (declarative knowledge): знания, представленные в виде фактов, правил и теорем.

Примечание – Обычно декларативные знания не могут быть обработаны без их предварительного преобразования в процедурные знания (3.1.28).

[ИСО/МЭК 2382–28:1995, 28.02.22]

3.1.13. **экспертная система** (expert system): ИИ-система (3.1.4), которая накапливает, комбинирует и объединяет знания (3.1.21), предоставленные людьми, являющимися экспертами в предметной области, с целью логического вывода решений для поставленных задач.

3.1.14. **универсальный ИИ, сильный ИИ** (general AI, AGI): тип ИИ-систем (3.1.4), решающих широкий круг задач (3.1.35) с приемлемым уровнем качества и производительности.

Примечание 1 – Ср. определение узконаправленного (слабого) искусственного интеллекта (3.1.24).

Примечание 2 – Термин «универсальный искусственный интеллект» часто используется для обозначения систем, которые могут выполнять не просто широкий спектр задач, а все задачи, которые способен выполнять человек.

3.1.15. **генетический алгоритм** (genetic algorithm, GA): алгоритм решения оптимизационных задач, имитирующий процесс естественного отбора посредством создания популяции особей (решений) и её последующей эволюции.

3.1.16. **гетерономия (гетерономность), гетерономный** (heteronomy, heteronomous): характеристика системы, функционирующей в условиях ограничений, связанных с внешним вмешательством, управлением или надзором.

3.1.17. **логический вывод** (inference): рассуждение, посредством которого из известных предпосылок делаются выводы.

Примечание 1 – В сфере ИИ предпосылкой может быть факт, правило, модель, признак либо необработанные данные.

Примечание 2 – Термин «логический вывод» относится как к процессу логического вывода, так и к его результату.

[ИСО/МЭК 2382:2015, 2123830]

3.1.18. **интернет вещей, IoT** (Internet of things, IoT): инфраструктура взаимосвязанных объектов, людей, систем и информационных ресурсов вместе с сервисами, которые обрабатывают и реагируют на информацию, поступающую из материального и виртуального миров.

[ИСО/МЭК 20924:2021, 3.2.4]

3.1.19. **устройство интернета вещей, IoT-устройство** (IoT device): объект IoT-системы (3.1.20), который взаимодействует и обменивается информацией с материальным миром посредством использования сенсоров (датчиков) и исполнительных устройств (приводов).

Примечание – IoT-устройство может быть датчиком и/или приводом.

[ИСО/МЭК 20924:2021, 3.2.6]

3.1.20. **система интернета вещей, IoT-система** (IoT system): система, обеспечивающая функциональные возможности интернета вещей (3.1.18).

Примечание – В состав IoT-системы могут входить (включая, но не ограничиваясь ими) IoT-устройства, IoT-шлюзы, сенсоры (датчики) и исполнительные устройства (приводы).

[ИСО/МЭК 20924:2021, 3.2.9]

3.1.21. **знания** (knowledge): <искусственный интеллект> абстрагированная информация об объектах, событиях, понятиях и правилах, их взаимосвязях и свойствах, организованная и упорядоченная для целенаправленного систематического использования.

Примечание 1 – В отличие от использования данного термина в некоторых других областях, в области ИИ «знания» не подразумевают наличия когнитивных способностей. В частности, «знания» не предполагают когнитивного акта понимания.

Примечание 2 – Информация может существовать в числовой и/или символьной форме.

Примечание 3 – Информация представляет собой данные, помещённые в определенный контекст (контекстуализированные), благодаря чему их можно интерпретировать. Данные создаются посредством абстрагирования или измерения явлений мира.

3.1.22. **жизненный цикл** (life cycle): эволюция системы, продукта, услуги, проекта или иной созданной человеком сущности от возникновения замысла до вывода из эксплуатации.

[ИСО/МЭК/ИИЭР 15288:2015, 4.1.23]

3.1.23. **модель** (model): физическое, математическое или иное представление системы, объекта, явления, процесса или данных.

[ИСО/МЭК 18023–1:2006, 3.1.11]

3.1.24. **узконаправленный искусственный интеллект, узконаправленный ИИ, слабый искусственный интеллект, слабый ИИ** (narrow AI): тип ИИ-систем (3.1.4), ориентированных на выполнение определённых задач (3.1.35) с целью решения конкретной проблемы.

Примечание – Ср. определение универсального (сильного) искусственного интеллекта (3.1.14).

3.1.25. **показатель (деятельности)** (performance): измеримый результат.

Примечание 1 – Под «показателями» могут пониматься как количественные, так и качественные результаты.

Примечание 2 – Показатели могут относиться к управленческой деятельности, процессам, продуктам (включая услуги), системам и/или организациям.

3.1.26. **планирование** (planning): <искусственный интеллект> вычислительные процессы, которые из набора действий формируют рабочий процесс, стремясь при этом к достижению определённой цели.

Примечание – Под «планированием», при использовании этого термина в стандартах жизненного цикла ИИ или менеджмента ИИ, также могут пониматься действия, предпринимаемые людьми.

3.1.27. **прогноз** (prediction): основной результат работы ИИ-системы (3.1.4), получаемый при подаче на вход системы входных данных (3.2.9) и/или информации.

Примечание 1 – Вслед за прогнозами могут быть получены дополнительные выходные результаты, такие как рекомендации, решения и действия.

Примечание 2 – Под «прогнозом» не обязательно понимается предсказание чего-либо в будущем.

Примечание 3 – Под «прогнозами» могут пониматься различные виды анализа и/или производства данных, применяемые к новым и/или историческим данным (включая перевод текста, создание синтетических изображений и диагностику последнего сбоя питания).

3.1.28. **процедурные знания** (procedural knowledge): знания, явным образом указывающие шаги, которые следует предпринять для решения задачи или достижения цели.

[ИСО/МЭК 2382–28:1995, 28.02.23]

3.1.29. **робот** (robot): оснащённая исполнительными устройствами (приводами) автоматическая система, которая выполняет целевые задачи (3.1.35) в материальном мире, измеряя с этой целью параметры своего окружения и используя программную систему управления.

Примечание 1 – В состав робота входят система управления и её интерфейс.

Примечание 2 – Робот классифицируется как промышленный либо сервисный в соответствии с его предполагаемым применением.

Примечание 3 – Чтобы надлежащим образом выполнять свои задачи (3.1.35), робот использует различные типы сенсоров (датчиков) для подтверждения своего текущего состояния и для восприятия элементов, образующих то окружение, в условиях которого он работает.

3.1.30. **робототехника** (robotics): теория и практика проектирования, производства и применения роботов.

[ИСО 8373:2012, 2.16]

3.1.31. **семантические вычисления** (semantic computing): область вычислений, стремящаяся определить смысл обрабатываемого контента (информации) и намерения пользователей, и представить их в машинно-обрабатываемой форме.

3.1.32. **мягкие вычисления** (soft computing): область вычислений, которые являются толерантными к неточностям, неопределенности и частичной истинности и используют их, чтобы сделать процесс решения задач более гибким и робастным.

Примечание – Понятие «мягкие вычисления» охватывают различные методы, такие как нечёткая логика, машинное обучение и вероятностные рассуждения.

3.1.33. **символьный искусственный интеллект, символьный ИИ, символический искусственный интеллект, символический ИИ** (symbolic AI): искусственный интеллект (3.1.3), основанный на методах и моделях (3.1.23), в которых для получения выводов используется манипулирование символами и структурами в соответствии с явно заданными правилами.

Примечание – Если сравнивать с субсимвольным ИИ (3.1.34), то символьный (символический) ИИ выдаёт логически выведенные декларативные результаты, в то время как субсимвольный ИИ основан на статистических подходах и выдает результаты с заданной вероятностью ошибки.

3.1.34. **субсимвольный искусственный интеллект, субсимвольный ИИ** (subsymbolic AI): искусственный интеллект (3.1.3), основанный на методах и моделях (3.1.23), использующих неявное кодирование информации, которое может быть выработано на основе опыта и/или необработанных (первоначальных) данных.

Примечание – Если сравнивать с символьным (символическим) ИИ (3.1.33), то субсимвольный ИИ основан на статистических подходах и выдает результаты с заданной вероятностью ошибки, в то время как символьный (символический) ИИ выдаёт декларативные результаты,

3.1.35. **задача** (task): <искусственный интеллект> действия, необходимые для достижения конкретной цели.

Примечание 1 – Действия могут быть физическими или когнитивными. Примерами могут служить вычисление или создание прогнозов (3.1.27), переводов, синтетических данных или артефактов; либо навигация в физическом пространстве.

Примечание 2 – Примерами задач являются классификация, регрессия, ранжирование, кластеризация и понижение размерности.

3.2 Термины, относящиеся к данным

3.2.1. **аннотирование данных, разметка данных** (data annotation): процесс присоединения к данным описательной информации, без внесения каких-либо изменений в сами данные.

Примечание – Описательная информация может принимать форму метаданных, меток и привязок.

3.2.2. **проверка качества данных** (data quality checking): процесс, в ходе которого данные проверяются на полноту, на предвзятость и на наличие иных факторов, влияющие на их полезность для ИИ-системы (3.1.4).

3.2.3. **аугментация данных** (data augmentation): процесс создания синтетических элементов данных посредством модификации существующих данных и/или выполнения операций над ними.

3.2.4. **выборка данных, процесс выборки данных** (data sampling): процесс формирования репрезентативного подмножества неделимых элементов данных, которое должно демонстрировать закономерности и тенденции, аналогичные тем, что свойственны анализируемому более объёмному набору данных (3.2.5).

Примечание – В идеале подмножество неделимых элементов данных должно быть репрезентативным по отношению к исходному, большему по объёму набору данных (3.2.5).

3.2.5. **набор данных, массив данных** (dataset): идентифицируемая совокупность данных, представленных в общем формате.

Примеры:

1 Сообщения в микроблоге за июнь 2020 года, помеченные хэштегами #регби и #футбол.

2 Макро-фотографии цветов в разрешении 256 x 256 пикселей.

Примечание – Наборы данных могут использоваться для валидации или тестирования ИИ-модели (3.1.23). В контексте машинного обучения (3.3.5) наборы

данных также могут использоваться для обучения алгоритма машинного обучения (3.3.6).

3.2.6. **разведочный анализ данных, исследовательский анализ данных, первичный анализ данных** (exploratory data analysis, EDA): первоначальное изучение данных для определения их явно выраженных характеристик и оценки их качества.

Примечание – Разведочный анализ данных может включать выявление отсутствующих значений и выбросов, определение репрезентативности для поставленной задачи – см. определение проверки качества данных (3.2.2).

3.2.7. **эталонное значение** (ground truth): значение целевой переменной, указанное для конкретного элемента размеченных входных данных.

Примечание – Данный термин не подразумевает последовательного соответствия указанных в размеченных входных данных «эталонных значений» значениям целевых переменных в реальном мире.

3.2.8. **подстановка недостающих значений** (imputation): процедура, в ходе которой недостающие данные заменяются данными, полученными в результате оценочных расчётов и/или моделирования.

[ИСО 20252:2019, 3.45]

3.2.9. **входные данные** (input data): данные, на основе которых ИИ-система (3.1.4) получает в качестве результата прогноз или логический вывод (3.1.17).

3.2.10. **метка** (label): значение целевой переменной, присвоенное неделимому элементу размеченных входных данных.

3.2.11. **персональные данные, ПДн** (personally identifiable information, PII, personal data): любая информация, которая (a) может быть использована для идентификации физического лица, к которому она относится, и/или (b) прямо или косвенно связана или может быть связана с физическим лицом.

Примечание 1 – «Физическое лицо» в данном определении является субъектом персональных данных. При установлении возможности идентифицировать субъекта ПДн следует принять во внимание все разумные средства, которые могут быть использованы для идентификации этого физического лица располагающим данными заинтересованным в ПДн лицом или любой иной стороной.

Примечание 2 – Данное определение включено для определения термина «персональные данные» в том смысле, в каком он используется в настоящем стандарте. Обработчик ПДн в публичном облаке, как правило, не может точно знать, относится ли обрабатываемая им информация к какой-либо конкретной категории, если только клиент облачных услуг не будет прозрачен в этом отношении.

[ИСО/МЭК 29100:2011/Amd1: 2018, 2.9]

3.2.12. **эксплуатационные данные, производственные данные** (production data): приобретённые на стадии эксплуатации ИИ-системы (3.1.4) данные, для которых развёрнутая ИИ-система (3.1.4) вычисляет в качестве результата прогноз или логический вывод (3.1.17).

3.2.13. **элемент данных** (sample): неделимый (в конкретном контексте) элемент данных; такие элементы в больших количествах обрабатываются алгоритмом машинного обучения (3.3.6) или ИИ-системой (3.1.4).

3.2.14. **тестовые данные** (test data, evaluation data): данные, используемые для оценки показателей работы окончательной модели (3.1.23).

Примечание – Тестовые данные не пересекаются с обучающими данными (3.3.16) и валидационными (проверочными) данными (3.2.15).

3.2.15. **валидационные данные, проверочные данные** (validation data, development data): данные, используемые для сравнения показателей работы различных моделей-кандидатов (3.1.23).

Примечание 1 – Валидационные (проверочные) данные не пересекаются с тестовыми данными (3.2.14) и, как правило, также и с обучающими данными (3.3.16). Однако в тех случаях, когда данных недостаточно для разделения их на три отдельных набора обучающих, валидационных и тестовых данных, данные разделяются только на два набора – тестовый набор данных и обучающий (либо валидационный) набор данных. Кросс-валидация и обобщённая кросс-валидация (bootstrapping) являются распространенными методами, используемыми для последующего создания отдельных наборов данных для обучения и валидации из обучающего (либо валидационного) набора данных.

Примечание 2 – Валидационные данные могут использоваться для настройки гиперпараметров и для валидации определенных алгоритмических решений, вплоть до решений о включении заданного правила в экспертную систему.

3.3 Термины, относящиеся к машинному обучению

3.3.1. **Байесовская сеть** (Bayesian network): вероятностная модель (3.1.23), использующая байесовский вывод (3.1.17) на направленном ациклическом графе для вычисления вероятности.

3.3.2. **дерево решений, дерево принятия решений (**decision tree): модель (3.1.23), логический вывод (3.1.17) для которой кодируется в виде путей от корня к листовой вершине в древовидной структуре.

3.3.3. **объединение человека и машины в команду** (human-machine teaming): интеграция способности человека к коллективному взаимодействию с возможностями машинного интеллекта.

3.3.4. **гиперпараметр** (hyperparameter): параметр алгоритма машинного обучения (3.3.6), влияющий на процесс обучения.

Примечание 1 – Гиперпараметры выбираются до начала обучения и могут использоваться в процессах для помощи в оценке параметров модели.

Примечание 2 – Примерами гиперпараметров могут служить количество слоев нейронной сети, ширина каждого слоя, тип функции активации, метод оптимизации, скорость обучения нейронных сетей; выбор функции ядра в методе опорных векторов; количество листьев или высота дерева; значение K при кластеризации методом K-средних; максимальное количество итераций алгоритма максимизации ожидания; количество гауссианов в гауссовой смеси.

3.3.5. **машинное обучение, МО** (machine learning, ML): процесс оптимизации параметров модели (3.3.8) с помощью вычислительных методов таким образом, чтобы поведение модели (3.1.23) отражало данные и/или опыт.

3.3.6. **алгоритм машинного обучения** (machine learning algorithm): алгоритм определения параметров (3.3.8) модели машинного обучения (3.3.7) в соответствии с заданными критериями на основе данных.

Пример – Рассмотрим задачу определения параметров линейной функции с одной переменной $y(x) = \theta_0 + \theta_1 x$***, где y – значение функции, x – независимая переменная,*** θ_0 ***– свободный член (значение функции при x = 0), и*** θ_1 ***– коэффициент. В машинном обучении (3.3.5) процесс определения свободного члена и коэффициентов линейной функции известен как линейная регрессия.***

3.3.7. **модель машинного обучения** (machine learning model): математическая конструкция, генерирующая логический вывод (3.1.17) или прогноз (3.1.27) на основе входных данных и/или информации.

Пример – По результатам обучения модели, представленной в виде линейной функции с одной переменной $y(x) = \theta_0 + \theta_1 x$, с использованием линейной регрессии, итоговая модель могла бы выглядеть как $y(x) = 3 + 7x$.

Примечание – Модель машинного обучения является результатом обучения на основе алгоритма машинного обучения (3.3.6).

3.3.8. **параметр, параметр модели** (parameter, model parameter): внутренняя переменная, влияющая на то, как модель (3.1.23) вычисляет свои выходные данные.

Примечание – Примерами параметров могут служить веса в нейронной сети и вероятности перехода в марковской модели.

3.3.9. **обучение с подкреплением** (reinforcement learning, RL): нахождение оптимальной последовательности действий для максимизации вознаграждения через взаимодействие с окружением, откликом которого являются сигналы подкрепления.

3.3.10. **повторное обучение, переобучение** (retraining): обновление обученной модели (3.3.14) посредством обучения (3.3.15) на иных обучающих данных (3.3.16).

3.3.11. **машинное обучение с частичным привлечением учителя, частично контролируемое обучение** (semi-supervised machine learning): машинное обучение (3.3.5), при котором в процессе обучения (3.3.15) используются как размеченные, так и неразмеченные данные.

3.3.12. **машинное обучение с учителем, контролируемое обучение** (supervised machine learning): машинное обучение (3.3.5), при котором в процессе обучения (3.3.15) используются только размеченные данные.

3.3.13. **метод опорных векторов, машина опорных векторов** (support vector machine, SVM): алгоритм машинного обучения (3.3.6), который максимизирует расстояние между границами решений.

Примечание – Опорные векторы представляют собой наборы точек данных, определяющие расположение границ решений (гиперплоскостей).

3.3.14. **обученная модель** (trained model): результат обучения модели (3.3.15).

3.3.15. **обучение, обучение модели** (training, model training): процесс определения или улучшения параметров модели машинного обучения (3.3.7) на основе алгоритма машинного обучения (3.2.10) с использованием обучающих данных (3.3.16).

3.3.16. **обучающие данные** (training data): данные, используемые для обучения модели машинного обучения (3.3.7).

3.3.17. **обучение без учителя, неконтролируемое обучение** (unsupervised machine learning): машинное обучение (3.3.5), при котором в процессе обучения (3.3.15) используются только неразмеченные данные.

3.4 Термины, относящиеся к нейронным сетям

3.4.1. **функция активации, активационная функция, передаточная функция** (activation function): функция, аргументом которой является взвешенная сумма входов нейрона (3.4.9).

Примечание – Функции активации позволяют нейронным сетям изучать сложные признаки в данных. Функции активации, как правило, нелинейны.

3.4.2. **свёрточная нейронная сеть, глубокая свёрточная нейронная сеть** (convolutional neural network, CNN, deep convolutional neural network, DCNN): нейронная сеть прямого распространения (3.4.6), использующая свёртку (3.4.3) по крайней мере в одном из своих слоёв.

3.4.3. **свёртка, конволюция** (convolution): математическая операция вычисления взаимной корреляции (скользящего скалярного произведения) входных данных.

3.4.4. **глубокое обучение (нейронной сети)** (deep learning, deep neural network learning): <искусственный интеллект> Подход к созданию обширных иерархических представлений посредством обучения (3.3.15) нейронных сетей (3.4.8) с большим количеством скрытых слоев.

Примечание – Глубокое обучение является частным случаем машинного обучения (3.3.5).

3.4.5. **взрывающийся градиент** (exploding gradient): явление, встречающееся при обучении (3.3.15) нейронных сетей с использованием алгоритма обратного распространения ошибок, когда начинают накаливаться большие значения градиента ошибок, приводящие к очень большим приращениям весовых коэффициентов, что делает модель (3.1.23) нестабильной.

3.4.6. **нейронная сеть прямого распространения, нейронная сеть с прямой связью** (feed forward neural network, FFNN): нейронная сеть (3.4.8), в которой информация передаётся только в одном направлении, от входного слоя к выходному.

3.4.7. **долгая краткосрочная память, длинная цепь элементов краткосрочной памяти** (long short-term memory, LSTM): тип рекуррентной нейронной сети (3.4.10), которая с приемлемой производительностью

обрабатывает последовательные данные как для коротких, так и для длинных интервалов последовательности.

3.4.8. **нейронная сеть, искусственная нейронная сеть** (neural network, NN, neural net, artificial neural network): <искусственный интеллект> сеть из двух или более слоёв, состоящих из нейронов (3.4.9), соединённых взвешенными связями с регулируемыми весовыми коэффициентами, при этом каждый нейрон получает входные данные и вырабатывает результат.

Примечание 1 – Нейронные сети являются ярким примером коннекционистского подхода (3.1.10).

Примечание 2 – Хотя первоначально источником идей для проектирования нейронных сетей послужило функционирование биологических нейронов, в настоящее время большинство работ по нейронным сетям уже не подвержено такому влиянию.

3.4.9. **нейрон** (neuron): <искусственный интеллект> базовый элемент, получающий одно или несколько входных значений и вырабатывающий выходное значение посредством комбинирования входных значений и применения функции активации (3.4.1) к результату комбинирования.

Примечание – Примерами нелинейных функций активации могут служить пороговая функция, сигмоидальная (сигмоидная) функция и полиномиальная функция.

3.4.10. **рекуррентная нейронная сеть** (recurrent neural network, RNN): нейронная сеть (3.4.8), в которой как выходные данные предыдущего слоя, так результаты предыдущего шага вычислений подаются на вход текущему слою.

3.5 Термины, относящиеся к надежности и доверию

3.5.1. **подотчётный** (accountable): обязанный отчитываться за действия, решения и показатели деятельности.

[ИСО/МЭК 29100:2011/Amd1: 2018, 2.9]

3.5.2. **подотчётность** (accountability): свойство быть подотчётным (3.5.1).

Примечание 1 – Подотчетность относится к ответственности, установленной для соответствующего лица или организации, которая может основываться на законах, нормативных правовых актах, соглашениях (контрактах) или же может быть установлена в рамках делегирования полномочий.

Примечание 2 – Подотчётность предполагает, что физическое или юридическое лицо должно отчитываться за что-либо перед другим физическим или юридическим лицом с использованием определённых средств и в соответствии с определёнными критериями.

[ИСО/МЭК 38500:2015, 2.3]

3.5.3. **доступность** (availability): свойство быть доступным и готовым к использованию по запросу авторизованного лица или устройства.

[ИСО/МЭК 27000:2018, 3.7]

3.5.4. **предвзятость, необъективность, смещённость** (bias): систематическое различие в отношении к определенным объектам, людям или группам по сравнению с другими.

Примечание – Под «отношением» здесь понимаются действия любого вида, включая восприятие, наблюдение, представление, прогноз (3.1.27) или принятие решения.

[ИСО/МЭК ТО 24027:2021, 3.3.2]

3.5.5. **управление, контроль и управление** (control): целенаправленное действие в рамках процесса или над ним для достижения определённых целей.

[МЭК 61800-7-1:2015, 3.2.6]

3.5.6. **управляемость, управляемый** (controllability, controllable): свойство ИИ-системы (3.1.4), означающее возможность человека или иного внешнего агента вмешиваться в функционирование системы.

3.5.7. **объяснимость** (explainability): свойство ИИ-системы (3.1.4) предоставлять информацию о влияющих на её результаты существенных факторах в понятном для людей виде.

Примечание – Цель объяснимости – дать ответ на вопрос «Почему?», не пытаясь при этом доказать, что выбранный вариант действий обязательно был оптимальным.

3.5.8. **предсказуемость** (predictability): свойство ИИ-системы (3.1.4), дающее возможность заинтересованным сторонам (3.5.13) делать надёжные предположения о результатах её работы.

[ИСО/МЭК ТО 27550:2019, 3.12]

3.5.9. **надёжность** (reliability): свойство последовательно демонстрировать ожидаемое поведение и результаты.

[ИСО/МЭК 27000:2018, 2.55]

3.5.10. **жизнеспособность, способность к восстановлению** (resilience): способность системы быстро восстанавливать рабочее состояние после инцидента.

3.5.11. **риск** (risk): влияние неопределенности на достижение целей.

Примечание 1 – Влияние выражается в отклонении от того, что ожидается. Оно может быть позитивным и/или негативным, и может приводить к реализации/устранению, созданию или появлению возможностей и угроз.

Примечание 2 – Цели могут иметь различные аспекты, относиться к различным категориям, и могут устанавливаться на различных уровнях.

Примечание 3 – Риски обычно описываются как сочетания источников риска, потенциальных событий, их вероятности и последствий.

[ИСО 31000:2018, 3.1]

3.5.12. **робастность** (robustness): способность системы поддерживать свой уровень показателей при любых обстоятельствах.

3.5.13. **заинтересованная сторона** (stakeholder): любое физическое лицо, группа или организация, которые могут повлиять на решение или действие, либо могут быть затронутыми или же посчитать себя затронутыми ими.

[ИСО/МЭК 38500:2015, 2.24]

3.5.14. **прозрачность** (transparency): <организация> характеристика организации, которая информирует соответствующие заинтересованные стороны (3.5.13) о затрагивающих их действиях и решениях всесторонним, доступным и понятным образом.

Примечание – Некорректное информирование о действиях и решениях может нарушить требования по безопасности, конфиденциальности и защите персональных данных.

3.5.15. **прозрачность** (transparency): <система> свойство системы, означающее, что надлежащая информация о системе предоставляется соответствующим заинтересованным сторонам (3.5.13).

Примечание 1 – С точки зрения обеспечения прозрачности системы, надлежащая информация может охватывать такие аспекты, как признаки, параметры производительности, ограничения, компоненты, процедуры, метрики, цели проектирования, проектные решения и допущения, источники данных и протоколы разметки (маркировки).

Примечание 2 – Некорректное раскрытие определённых аспектов системы может нарушить требования по безопасности, конфиденциальности и защите персональных данных.

3.5.16. **свойство вызывать доверие, надёжность** (trustworthiness): способность проверяемым образом удовлетворять ожидания заинтересованных сторон (3.5.13).

Примечание 1 – В зависимости от контекста (условий деятельности) или сектора, а также от конкретного продукта или услуги (сервиса), от используемых данных и технологий, - важными являются различные аспекты доверия, верификация которых требуется для обеспечения того, что ожидания заинтересованных сторон (3.5.13) удовлетворяются.

Примечание 2 – Цели могут иметь различные аспекты, относиться к различным категориям, и могут устанавливаться на различных уровнях В число аспектов доверия входят, например, надежность, доступность, жизнеспособность, защищённость, неприкосновенность частной жизни (защита персональных данных), безопасность, подотчетность, прозрачность, целостность, аутентичность, качество и пригодность к использованию (удобство использования).

Примечание 3 – О свойстве вызывать доверие (надёжности) можно говорить в отношении услуг (сервисов), продуктов, технологий, данных и информации, а также, в контексте стратегического управления, в отношении организаций.

[ИСО/МЭК ТО 24028:2020, 3.42]

3.5.17. **верификация** (verification): подтверждение, посредством представления объективных доказательств, того, что установленные требования были выполнены.

Примечание – Верификация обеспечивает уверенность только лишь в том, что продукт соответствует своим спецификациям.

[ИСО/МЭК 27042:2015, 3.2]

3.5.18. валидация (validation): подтверждение, посредством представления объективных свидетельств того, что требования в отношении конкретного предполагаемого использования или применения были выполнены.

[ИСО/МЭК 27043:2015, 3.16]

3.6 Термины, относящиеся к обработке естественного языка

3.6.1. **автоматическое реферирование** (automatic summarization): задача (3.1.35) сокращённого изложения контента или текста на естественном языке (3.6.7) при сохранении важной семантической информации.

3.6.2. **управление диалогом** (dialogue management): задача (3.1.35) выбора подходящего следующего шага в диалоге на основе пользовательского ввода, истории диалога и других контекстуальных знаний (3.1.21), в интересах достижения желаемой цели.

3.6.3. **распознавание эмоций** (emotion recognition): задача (3.1.35) компьютерной идентификации и классификации эмоций, выраженных в фрагменте текста, в речи, на видео, в изображении или в их комбинации.

Примечание – Примерами эмоций могут служить счастье, печаль, гнев и восторг.

3.6.4. **извлечение информации** (information retrieval, IR): задача (3.1.35) извлечения из набора данных (3.2.5) релевантных материалов или их частей, обычно на основе запросов по ключевым словам или запросов на естественном языке (3.6.7).

3.6.5. **машинный перевод** (machine translation, MT): задача (3.1.35) автоматического перевода текста или речи с одного естественного языка (3.6.7) на другой с помощью компьютерной системы.

[ИСО 17100:2015, 2.2.2]

3.6.6. **распознавание именованных сущностей** (named entity recognition, NER): задача (3.1.35) распознавания и разметки денотативных (понимаемых буквально) наименований сущностей и их категорий для последовательностей слов в потоке текста или речи.

Примечание 1 – Под сущностями понимаются представляющие интерес конкретные или абстрактные вещи (объекты), включая ассоциации между вещами.

Примечание 2 – Под «поименованной сущностью» понимается сущность с денотативным наименованием, имеющим конкретное или уникальное значение.

Примечание 3 – К денотативным наименованиям относятся имена конкретных лиц, мест, организаций и иные имена собственные, в зависимости от предметной области или приложения.

3.6.7. **естественный язык** (natural language): язык, который активно используется или ранее активно использовался сообществом людей, правила которого обусловлены практикой его применения.

Примечание 1 – Естественным языком является любой человеческий язык, который может быть выражен в виде текста, речи, языка жестов и т.д.

Примечание 2 – Естественным языком является любой язык общения между людьми, такой как русский, английский, испанский, арабский, китайский или японский языки. Естественные языки следует отличать от языков программирования и формальных языков, таких как Java, Fortran, C++ или логика (исчисление предикатов) первого порядка.

[ИСО/МЭК 15944-8:2012, 3.82]

3.6.8. **генерация естественного языка** (natural language generation, NLG): задача (3.1.35) преобразования несущих семантику данных в естественный язык (3.6.7).

3.6.9. **обработка естественного языка** (natural language processing, NLP): <система> обработка информации на основе понимания естественного языка (3.6.11) и/или генерация естественного языка (3.6.8).

3.6.10. **обработка естественного языка** (natural language processing, NLP): <дисциплина> дисциплина, изучающая то, как системы воспринимают, обрабатывают и интерпретируют естественный язык (3.6.7).

3.6.11. **понимание естественного языка** (natural language understanding, NLU, natural language comprehension): извлечение функциональным компонентом информации из текста или речи, переданных ему на естественном языке (3.6.7), и создание описания как этого текста или речи, так и того, что они представляют.

[ИСО/МЭК 2382:2015, 2123786]

3.6.12. **оптическое распознавание символов** (optical character recognition, OCR): преобразование изображений машинописного, печатного или рукописного текста в машиночитаемый текст.

3.6.13. **морфологическая разметка, частеречная разметка** (part-of-speech tagging): задача (3.1.35) присвоения слову категории (такой, например, как глагол, существительное, прилагательное) на основе его грамматических свойств.

3.6.14. **поиск ответа на вопрос** (question answering): задача (3.1.35) определения наиболее подходящего ответа на вопрос, заданный на естественном языке (3.6.7).

Примечание – Вопрос может предполагать как выбор предполагаемого ответа из списка, так и ответ в свободной форме.

3.6.15. **извлечение взаимосвязей** (relationship extraction, relation extraction): задача (3.1.35) выявления отношений между упомянутыми в тексте сущностями.

3.6.16. **анализ тональности, анализ настроений, анализ эмоциональной окраски, сентимент-анализ** (sentiment analysis): задача (3.1.35) выявления и категоризации с помощью вычислительных методов мнений, выраженных во фрагменте текста, речи или изображения, с целью определения характера эмоций или отношения, например, в диапазоне от позитивного до негативного.

Примечание – Примерами тональности (эмоциональной окраски) могут служить одобрение и неодобрение, позитивное и негативное отношение, согласие и несогласие.

3.6.17. **распознавание речи** (speech recognition, speech-to-text, STT): преобразование функциональным компонентом речевого сигнала в представление содержания речи.

[ИСО/МЭК 2382:2015, 2120735]

3.6.18. **синтез речи** (speech synthesis, text-to-speech, TTS): генерация искусственной речи.

[ИСО/МЭК 2382: 2015, 2120745]

3.7 Термины, относящиеся к компьютерному зрению

3.7.1. **компьютерное зрение, машинное зрение** (computer vision): способность функционального компонента получать, обрабатывать и интерпретировать данные, представляющие изображения или видеосигналы.

Примечание – Компьютерное зрение включает использование датчиков (сенсоров) для создания цифрового образа визуальной сцены. Оно может включать обработку изображений, полученных в диапазонах длин волн, находящихся вне диапазона длин волн видимого света, таких как изображения в инфракрасных лучах.

3.7.2. **распознавание лиц** (face recognition): процесс автоматического распознавания образов, сравнивающий изображение реального лица с сохранёнными изображениями, отмечая при этом совпадения (если они есть) и выдавая сведения о личности идентифицируемого лица.

[ИСО 5127:2017, 3.1.12.09]

3.7.3. **изображение, образ** (image): <цифровые технологии> графический контент, предназначенный для визуального представления.

Примечание – Данное понятие охватывает графические материалы, закодированные в любом из электронных форматов, включая (но не ограничиваясь ими) растровые изображения, состоящие из отдельных пикселей (например, созданные программами для рисования или фотографическими средствами), и изображения, закодированные в виде набора формул (например, созданные в формате масштабируемой векторной графики).

[ИСО/МЭК 20071-1- 2019, 3.2.1]

3.7.4. **распознавание образов, распознавание изображений** (image recognition): процесс классификации объектов, типовых элементов и/или их конфигураций, представленных на изображении (3.7.3).

4 Сокращения

API — интерфейс программирования приложений (application programming interface)

КФС — киберфизическая система (cyber-physical system, CPS)

CPU — центральный процессор (central processing unit)

CRISP-DM — межотраслевой стандартный процесс интеллектуального анализа данных (cross-industry standard process for data mining)

DNN — глубокая нейронная сеть (deep neural network)

ЦСП — цифровой сигнальный процессор (digital signal processor, DSP)

FPGA — программируемая логическая интегральная схема (field-programmable gate array)

GPU — графический процессор (graphics processing unit)

СММ — скрытая марковская модель (hidden Markov model, HMM)

ИТ — информационные технологии

NPU — нейронный процессор, аппаратный ускоритель для нейронных сетей (neural processing unit)

ОЭСР — Организация экономического сотрудничества и развития (Organization for Economic Co-operation and Development, OECD)

POS — морфологический, частеречный (part of speech)

МО — машинное обучение

5 Понятия искусственного интеллекта

5.1 Общие положения

Междисциплинарные исследования и проекты разработки ИИ-систем направлены на создание компьютерных систем, способных выполнять задачи, которые обычно требуют интеллекта. Машины, использующие ИИ, предназначены для восприятия определенных сред и для выполнения действий, удовлетворяющих потребности этих машин.

Искусственный интеллект использует методы из многих областей знаний, таких как информатика, математика, философия, лингвистика, экономика, психология и когнитивные науки.

По сравнению с большинством традиционных информационных систем, не имеющих искусственного интеллекта, существует ряд интересных особенностей, общих для всех или некоторых ИИ-систем:

a) Интерактивность - входные данные ИИ-систем генерируются датчиками (сенсорами) и/или посредством взаимодействия с людьми; а их выходные данные могут привести к подаче управляющего сигнала на исполнительные устройства или к выдаче ответов людям или машинам. Примером может служить распознавание объекта в результате представления ИИ-системе его изображения.

b) Контекстуальность - некоторые ИИ-системы могут использовать несколько источников информации, включая источники как структурированной, так и неструктурированной цифровой информации, а также данные, получаемые от датчиков.

c) Надзор со стороны человека – ИИ-системы могут функционировать при различной степени человеческого надзора и контроля, зависящей от области применения. Примером могут служить самоуправляемые автомобили с разными уровнями автоматизации.

d) Адаптивность - некоторые ИИ-системы проектируются таким образом, чтобы использовать поступающие в режиме реального времени динамические данные и переобучаться, модифицируя на основе новых данных свой способ работы.

5.2 От сильного и слабого искусственного интеллекта к универсальному и узконаправленному

В своё время возможность создания обладающих интеллектом машин активно обсуждалась с философской точки зрения. Эти дискуссии привели к выделению двух разных видов ИИ: так называемых «слабого» и «сильного» ИИ. В случае слабого ИИ, ИИ-система может лишь обрабатывать символы (буквы, цифры и т.д.), даже не понимая, что именно она делает. В случае «сильного» ИИ, ИИ-система тоже обрабатывает символы, но при этом она по-настоящему «понимает», что

делает. Категории «слабый ИИ» и «сильный ИИ» в основном важны для философов, но не актуальны для исследователей и практиков в сфере искусственного интеллекта.

Позднее появились противопоставляемые друг другу категории «узконаправленный ИИ» и «универсальный ИИ», которые больше подходят для сферы искусственного интеллекта. Система «узконаправленного ИИ» способна выполнять определенные задачи для решения конкретной проблемы (возможно, намного лучше, чем это сделали бы люди). Система «универсального ИИ» способна выполнять широкий спектр задач с приемлемым уровнем эффективности и производительности. Современные ИИ-системы считаются системами «узконаправленного ИИ». Пока ещё неясно, будут ли ИИ-системы «универсального ИИ» технически осуществимыми в будущем.

5.3 ИИ-система как агент

Поскольку некоторые приложения ИИ нацелены на моделирование человеческого интеллекта и человеческого поведения, на ИИ-системы можно смотреть с точки зрения парадигмы действующего лица-агента. С инженерной точки зрения искусственный интеллект можно рассматривать как прикладную область, стремящуюся создать искусственных агентов, демонстрирующих рациональное поведение. В парадигме агента проводится чёткое разграничение между агентом и тем окружением, в условиях которого он эволюционирует. Данная парадигма проиллюстрирована на рисунке 1.

Рисунок 1 — Парадигма агента

ИИ-агент взаимодействует со своим окружением посредством датчиков (сенсоров) и исполнительных устройств (приводов), предпринимая действия, которые максимизируют его шансы на успешное достижение своих целей.

В зависимости от выполняемой задачи, окружения могут иметь различные характеристики, которые влияют на уровень сложности решения проблемы.

В рамках данной парадигме можно выделить несколько типов ИИ-агентов в зависимости от их архитектуры [29]:

- рефлекторно действующие агенты, которые при выборе действия полагаются только на текущую ситуацию;

- агенты, действующие на основе моделей, которые полагаются на модель своего окружения, что позволяет им учитывать ожидаемые результаты доступных для них действий;

- целеориентированные (ориентированные на полезность) агенты, которые полагаются на внутреннюю функцию полезности, позволяющую им выбирать действия, ведущие к достижению целей, а среди целей выбирать наиболее привлекательные;

- обучающиеся агенты, способные собирать информацию о своём окружении и обучаться с целью улучшения показателей своей деятельности.

5.4 Знания

Специфическая для сферы искусственного интеллекта интерпретация понятия «знания» заслуживает более подробного обсуждения ввиду частого употребления этого понятия как в настоящем стандарте, так и в предметной области.

Если в других областях знаний данный термин может ассоциироваться с когнитивными способностями, то в контексте ИИ - это чисто технический термин, который относится к содержанию, а не к способностям и возможностям. Понятие «знаний» является частью иерархии «данные – информация – знания», согласно которой данные могут использоваться для производства информации, а информация может использоваться для производства знаний. В контексте ИИ это чисто технические, некогнитивные процессы.

Знания отличаются от информации тем, что информация наблюдается системой, а знания - это то, что система сохраняет по итогам таких наблюдений. Знания структурированы и организованы; они абстрагируются от особенностей отдельных наблюдений. В зависимости от цели, одна и та же информация может привести к разным знаниям.

Знания отличаются от их представления в том, что одни и те же знания могут иметь различные представления: они могут принимать разные конкретные формы, пригодные для передачи или хранения, каждая из которых имеет свои достоинства и недостатки, но при этом смысл у всех у них один и тот же.

Эти различия имеют техническое значение, поскольку некоторые подходы, методы и другие аспекты изучения ИИ полностью опираются на способность создавать разные знания на основе одной и той же информации или разные представления одних и тех же знаний.

5.5 Процесс познания и когнитивные вычисления

Познание включает в себя приобретение и обработку знаний посредством рассуждений, индивидуального или коллективного опыта, обучения и восприятия. Оно охватывает такие понятия, как внимание, формирование знаний, память, суждение и оценка, рассуждение и вычисления, решение задач и принятие решений, понимание и порождение речи.

Когнитивные вычисления являются одной из дисциплин, составляющих ИИ [27]. Они направлены на реализацию процесса познания с использованием таких возможностей, как выявление закономерностей при обработке огромных объемов данных. Когнитивные вычисления позволяют людям более естественно взаимодействовать с машинами. Задачи когнитивных вычислений связаны с машинным обучением, обработкой речи, обработкой естественного языка, компьютерным зрением и человеко-машинными интерфейсами.

5.6 Семантические вычисления

Семантические вычисления направлены на сопоставление семантики обрабатываемого контента с человеческими намерениями. В ходе семантических вычислений создаются представления для описания информации, которые используются для извлечения, управления,

манипулирования и создания контента (например, текста, видео, аудио, процесса, функции, устройства и сети). Семантическое описание контента позволяет уменьшить неопределенность в когнитивных процессах и в логических рассуждениях на основе информации. Это, в свою очередь, помогает обеспечить обогащение информации, устранение конфликтов, реферирование и сравнение. Таким образом, семантические вычисления - это подход, который сочетает в себе использование априорной информации и обучения.

5.7 Мягкие вычисления

Мягкие вычисления - это методология, которая сочетает в себе различные методы, толерантные к неточностям, неопределенности и частичной истинности при решении сложных задач. Традиционные вычислительные методы обычно применяются для поиска точных и строгих решений задач. Такие решения, однако, могут оказаться неподходящими или, как альтернатива, чрезвычайной сложными. Мягкие вычисления опираются на понимание того, что реальный мир часто неточен и неопределён, ввиду чего попытки отыскать точные решения реальных проблем часто могут быть сопряжены с затратами и сложностью. Таким образом, мягкие вычисления связаны с толерантным отношением к неточностям, неопределенности и частичной истинности для достижения гибких, робастных (устойчивых) и недорогих решений [24]. Примерами методов мягких вычислений являются нечёткие системы, эволюционные алгоритмы, роевой интеллект и системы на основе нейронных сетей.

5.8 Генетические алгоритмы

Генетические алгоритмы имитируют процесс естественного отбора, создавая и осуществляя эволюцию популяции особей (решений) в задачах оптимизации. Метод создания новых решений на основе исходной популяции имитирует генетические мутации. Хромосома (набор «генов») представлена в виде цепочки нулей и единиц. После создания исходной популяции хромосом первым шагом является вычисление приспособленности (пригодности) каждой хромосомы. Значение функции приспособленности является количественной оценкой оптимальности решения, ранжируя его по сравнению с другими решениями. Если созданное решение не является оптимальным, то выбирается пара хромосом, которые обмениваются своими частями (кроссовер), создавая двух хромосом-потомков. На следующем шаге выполняется мутация - случайным образом изменяется как минимум один ген в хромосомах. Исходная популяция заменяется новой популяцией, и начинается новая итерация. Итерации генетического алгоритма заканчиваются, когда выполняется один из критериев завершения (обычно достижение предопределенного числа итераций). В конечном итоге сохраняются наиболее приспособленные хромосомы [25].

5.9 Символьный и субсимвольный подходы к ИИ

В дисциплине искусственного интеллекта существует много разных точек зрения с различными парадигмами. Не существует единственной классификации, которая бы установила чёткое различие между разными типами ИИ. Тем не менее, можно указать ряд направлений, по которым можно позиционировать ИИ-системы.

С момента основания искусственного интеллекта как дисциплины в конкуренции друг с другом развивались две парадигмы: символьный ИИ и субсимвольный ИИ.

Символьный ИИ предполагает кодирование знаний с помощью символов и структур, когда для моделирования процессов рассуждений в основном используется логика. В рамках данной парадигмы информация явно кодируется с использованием формального представления, синтаксис которого может обрабатываться компьютером, а семантика имеет смысл для человека.

Второй подход - это субсимвольный ИИ, использующий коннекционистскую парадигму. Данная парадигма опирается на неявное кодирование знаний, а не на рассуждения, осуществляемые посредством манипулирования символами. Это неявное представление знаний преимущественно основано на статистических подходах к обработке опыта и/или необработанных (первоначальных) данных. Примерами ИИ-систем этого типа являются различные системы машинного обучения, включая различные виды глубоких нейронных сетей.

Современные ИИ-системы обычно содержат элементы как символьного, так и субсимвольного ИИ. Такие системы называются гибридным ИИ.

5.10 Данные

Данные играют центральную роль во многих ИИ-системах. Многие из этих систем спроектированы для оперирования данными, и часто бывает необходимо использовать данные для целей тестирования. В случае систем машинного обучения весь их жизненный цикл зависит от наличия и доступности данных.

Данные могут поступать как в структурированной форме (например, в виде реляционных баз данных), так и в неструктурированной форме (например, сообщения электронной почты, текстовые документы, изображения, аудио- и видеофайлы). Данные являются ключевым аспектом ИИ-систем, проходя через различные процессы, включающие, в том числе, следующие:

- комплектование данных, при котором данные получаются из одного или нескольких источников. Данные могут собраны из внутренних источников для оперирования данными для оперирования данными организации или же могут поступить извне. Необходимо оценить пригодность данных - например, определить, не являются ли они в той или иной степени предвзятыми или необъективными; и являются ли они достаточно обширными, чтобы адекватно представлять ожидаемые эксплуатационные входные данные;

- разведочный (первичный) анализ данных, при котором данные изучаются на предмет наличия в них закономерностей, взаимосвязей, тенденций и выбросов. Результаты такого анализа могут направлять дальнейшие шаги, такие как обучение и верификация;

- аннотирование данных, в ходе которой существенные элементы данных добавляются в качестве метаданных (например, информация о происхождении данных или же метки, которые помогают обучать модель);

- подготовка данных, в ходе которой данные преобразуются в форму, которую может использовать ИИ-система;

- фильтрация, представляющая собой удаление нежелательных данных. Эффекты от фильтрации необходимо тщательно изучить, чтобы избежать внесения нежелательной систематической ошибки (предвзятости) и возникновения иных проблем;

- нормализация, представляющая собой приведение значений данных к единому масштабу с тем, чтобы они были математически сопоставимы;

- обезличивание или иные процессы, проведение которых может потребоваться в том случае, если набор данных включает персональные данные или же ассоциирован с отдельными лицами или организациями, - прежде чем эти данные могут быть использованы ИИ-системой (см., например, стандарт ISO/IEC 20889);

- проверка качества данных, в рамках которой содержание данных изучается на предмет полноты, предвзятости и иных факторов, влияющих на полезность данных для ИИ-системы. Проверка на отравление (порчу) данных играет ключевую роль для обеспечения того, чтобы обучающие данные не были загрязнены данными, способными привести к вредным или нежелательным результатам;

- формирование выборки данных, когда извлекается репрезентативное подмножество данных;

- аугментация данных, при которой те данные, что имеются в слишком малых количествах, подвергаются нескольким видам преобразований с целью расширения набора данных;

- разметка данных, при которой наборы данных снабжаются метками, что означает, что неделимые элементы данных связываются со значениями целевых переменных. Метки часто необходимы для тестовых и валидационных данных. Некоторые подходы машинного обучения также основываются на наличии меток при обучении модели (см. 5.11.1 и 5.11.3).

В зависимости от варианта использования и применяемого подхода, данные в ИИ-системе могут быть задействованы несколькими способами:

- эксплуатационные (производственные) данные – это данные, получаемые и обрабатываемые ИИ-системой на стадии эксплуатации. В

зависимости от варианта использования, не все ИИ-системы используют эксплуатационные данные, однако это не зависит от технического проектного решения ИИ-системы и применяемого подхода.

- тестовые данные – это данные, используемые для оценки показателей работы ИИ-системы до её развертывания. Ожидается, что тестовые данные будут схожи с эксплуатационными данными, и для корректной оценки ИИ-системы необходимо, чтобы тестовые данные не пересекались с какими-либо иными данными, используемыми в процессе разработки. Проведение оценки требуется при использовании любого из методов и подходов ИИ, однако, в зависимости от задачи, использовать для этой цели тестовые данные не всегда уместно.

- валидационные данные – это данные, используемые разработчиком для принятия или проверки некоторых алгоритмических решений (таких, как подбор гиперпараметров, разработку правил и т.д.). Эти данные носят разные названия в зависимости от предметной области ИИ; например, при обработке естественного языка их обычно называют «данными разработки» (development data). Бывают ситуации, в которых валидационные данные не нужны;

- обучающие данные используются в специфическом контексте машинного обучения: они служат тем исходным материалом, из которого алгоритм машинного обучения «извлекает» свою модель для решения поставленной задачи.

Примечание 1 – В концептуальных структурах оценки программного обеспечения валидация - это процесс проверки выполнения определённых требований. Он является частью процесса оценки. В контексте ИИ термин «валидация» используется для обозначения процесса использования данных для выбора определённых значений и свойств, относящихся к проектному решению ИИ-системы. Здесь речь не идёт об оценке системы с точки зрения предъявляемых к ней требований, и всё это происходит до стадии оценки.

Примечание 2 – В концептуальных структурах оценки программного обеспечения под тестированием могут пониматься различные процессы, такие как поиск ошибок, выполнение тестов функциональных модулей и измерение времени вычислений. В области ИИ данный термин относится конкретно к статистической оценке параметров работы системы с использованием специального набора данных.

5.11 Понятия машинного обучения

5.11.1 Машинное обучение с учителем

Машинное обучение с учителем (контролируемое обучение) определяется как «машинное обучение, при котором в процессе обучения используются только размеченные данные» (3.3.12). В этом случае модели машинного обучения обучаются с помощью обучающих данных, которые включают в себя известное или определённое значение результата или целевой переменной (метку). Значение целевой переменной для данного неделимого элемента данных также известно как «эталонное значение». Метки могут быть любого типа, включая, в зависимости от задачи, категориальные, двоичные или числовые значения, или же структурированные объекты (например, последовательности, изображения, деревья или графы). Метки могут быть частью исходного набора данных, однако во многих случаях они определяются вручную или с помощью других процессов.

Обучение с учителем можно использовать для задач классификации и регрессии, а также для более сложных задач, связанных со структурированным прогнозированием.

Дополнительную информацию о машинном обучении с учителем см. в стандарте ИСО/МЭК 23053.

5.11.2 Машинное обучение без учителя

Машинное обучение без учителя (неконтролируемое обучение) определяется как «машинное обучение, при котором в процессе обучения используются только неразмеченные данные» (3.3.17).

Машинное обучение без учителя может быть полезно в таких случаях, как проведение кластерного анализа (кластеризации), когда поставлена задача выявления общих черт у неделимых элементов данных в составе входных данных. Ещё одним применением машинного обучения без учителя является уменьшение размерности обучающего набора данных, когда наиболее статистически значимые признаки определяются вне зависимости от наличия каких-либо меток.

Дополнительную информацию о машинном обучении без учителя см. в стандарте ИСО/МЭК 23053.

5.11.3 Машинное обучение с частичным привлечением учителя

Машинное обучение с частичным привлечением учителя (частично контролируемое обучение) определяется как «машинное обучение, при котором в процессе обучения используются как размеченные, так и неразмеченные данные» (3.3.11). Такой вид машинного обучения представляет собой гибрид машинного обучения с учителем и без учителя.

Машинное обучение с частичным привлечением учителя полезно в тех случаях, когда разметка всех неделимых элементов данных в большом наборе обучающих данных была бы непосильной с точки зрения времени и/или затрат. Дополнительные сведения о машинном обучении с частичным привлечением учителя см. в стандарте ИСО/МЭК 23053.

5.11.4 Обучение с подкреплением

Обучение с подкреплением – это процесс обучения агента (агентов), взаимодействующего со своим окружением ради достижения заранее определённой цели. В ходе обучения с подкреплением агент машинного обучения обучается посредством итеративного процесса проб и ошибок. Цель агента заключается в поиске стратегии (т.е. построении модели) для получения от окружения наилучшего вознаграждения. При каждой попытке (успешной или неуспешной) окружение предоставляет косвенную обратную связь. Затем агент корректирует своё поведение (т.е. свою модель) на основе этой обратной связи. Дополнительную информацию об обучении с подкреплением можно найти в стандарте ИСО/МЭК 23053.

5.11.5 Трансферное обучение

Трансферное обучение («перенос обучения») относится к серии методов, в которых знания, полученные на основе данных, предназначенных для решения одной проблемы, используются для решения иной проблемы. Например, информацию, полученную при распознавании номеров домов при просмотре изображений улиц, можно использовать для распознавания рукописных чисел. Дополнительную информацию о трансферном обучении можно найти в стандарте ИСО/МЭК 23053.

5.11.6 Обучающие данные

Обучающие данные состоят из неделимых элементов данных, используемых для обучения алгоритма машинного обучения. Как правило, эти неделимые элементы данных относятся к какому-либо конкретному интересующему вопросу и могут состоять из

структурированных или неструктурированных данных. Неделимые элементы данных могут быть неразмеченными и размеченными.

В последнем случае метка используется для управления процессом обучения модели машинного обучения. Например, если входными данными являются изображения и цель заключается в том, чтобы решить, показывает ли изображение кошку, то значением метки может быть «истина» для изображений с кошкой и «ложь» для изображений, на которых кошки нет. Это позволяет обученной модели отразить статистическую взаимосвязь между атрибутами неделимых элементов обучающих данных и значениями целевой переменной.

Количество неделимых элементов данных в наборе обучающих данных и выбор соответствующих признаков влияют на то, насколько хорошо результирующая модель машинного обучения соответствует распределению данных или значений целевой переменной. Однако в случае, если набор данных чрезвычайно велик, приходится идти на компромисс с учётом времени вычислений и необходимых для вычислений ресурсов.

5.11.7 Обученная модель

В настоящем стандарте обученная модель определяется как результат процесса обучения модели, который, в свою очередь, понимается как определение или улучшение параметров модели машинного обучения на основе алгоритма машинного обучения с использованием обучающих данных. Модель машинного обучения - это математическая конструкция, порождающая логический вывод или прогноз на основе входных данных и/или информации. Обученная модель должна быть пригодной для использования ИИ-системой при получении прогноза на основе эксплуатационных данных из интересующей области. Существуют различные стандартизированные

форматы для хранения и передачи обученных моделей в виде набора чисел.

5.11.8 Валидационные и тестовые данные

Для проведения оценки обученной модели обычной практикой является разделение данных, приобретённых для разработки модели, на наборы данных для обучения, валидации и тестирования.

Валидационные данные используются в ходе и после обучения для настройки гиперпараметров. Тестовые данные используются для проверки того, что модель научилась тому, чему она должна была научиться. Оба этих набора состоят из данных, которые никогда не показываются модели во время обучения. Если же для этих целей использовать обучающие данные, то модель способна «запомнить» правильный прогноз без фактической обработки выборки данных. Во избежание завышенной оценки показателей производительности модели тестовые данные также не показываются модели во время её настройки.

При использовании перекрестной проверки данные разделяются таким образом, что каждый неделимый элемент данных использовался как для обучения, так и для валидации. Такой подход имитирует использование большего по объёму набора данных, что может повысить показатели производительности модели. Иногда имеющихся данных бывает недостаточно для того, чтобы можно было сформировать отдельные наборы данных для обучения, валидации и тестирования. В таких случаях данные разбиваются только на два набора, а именно: 1) обучающие/валидационные данные и 2) тестовые данные. Затем на основе обучающих/валидационных данных генерируются отдельные наборы валидационных данных и обучающих данных - например, с помощью кросс-валидации и обобщённой кросс-валидация (boot-strapping).

5.11.9 Повторное обучение

5.11.9.1. Общие положения

Повторное обучение (переобучение) состоит в обновлении обученной модели посредством обучения на иных обучающих данных. Такая необходимость может возникнуть по многим причинам, включая отсутствие больших обучающих наборов данных, дрейф данных и дрейф концепции.

При дрейфе данных точность вычисляемых моделью прогнозов со временем снижается из-за изменений в статистических характеристиках эксплуатационных данных (например, может измениться разрешение изображений; или один класс может начать чаще встречаться в данных, чем другой). В этом случае модель необходимо переобучить на новых обучающих данных, которые лучше отражают текущие эксплуатационные данные.

При дрейфе концепции смещается граница принятия решений (например, представление о том, что является законным, а что нет, имеет тенденцию меняться после публикации новых законов), что также снижает точность прогнозов, даже если сами данные не изменились. В случае дрейфа концепции значения целевых переменных в обучающих данных необходимо переразметить, а модель - переобучить.

При повторном обучении существующей модели особое внимание уделяется преодолению или минимизации проблем, связанных с так называемым «катастрофическим забыванием». Многие алгоритмы машинного обучения хорошо справляются с задачами обучения только если данные представлены все сразу. По мере обучения модели для решения конкретной задачи её параметры адаптируются для решения данной задачи. Когда вводятся новые обучающие данные, то адаптации, основанные на этих новых наблюдениях «перезаписывают» знания,

которые модель приобрела ранее. Для нейронных сетей это явление известно как «катастрофическое забывание», и оно считается одним из их фундаментальных ограничений.

5.11.9.2. Непрерывное обучение

Непрерывное обучение, также известное как инкрементальное (продолжающееся) обучение или обучение на протяжении всего жизненного цикла, представляет собой последовательное обучение модели, которое продолжается на постоянной основе на всей стадии эксплуатации в жизненном цикле ИИ-системы. Это частный случай повторного обучения, когда обновления модели повторяются с высокой частотой и не влекут за собой прерывание работы.

Во многих ИИ-системах система обучается в процессе разработки, до её ввода в промышленную эксплуатацию. По своему характеру это похоже на стандартную разработку программного обеспечения, когда система создается и полностью тестируется перед вводом в эксплуатацию. Поведение таких систем оценивается в ходе процесса верификации, и ожидается, что оно не изменится на стадии эксплуатации.

Для ИИ-систем, использующих непрерывное обучение, характерно постепенное инкрементальное обновление модели в системе по мере её работы на стадии эксплуатации. Данные, вводимые в систему во время работы, не только анализируются с целью получения от системы результата, но и одновременно используются для настройки модели в системе с целью её улучшения на основе эксплуатационных данных. В зависимости от архитектуры ИИ-системы с непрерывным обучением, по ходу этого процесса могут потребоваться действия человека, такие, например, как разметка данных, валидация результатов определенного инкрементального обновления или же мониторинг показателей производительности ИИ-системы.

Непрерывное обучение может помочь справиться с последствиями ограниченности первоначальных обучающих данных, а также с дрейфом данных и дрейфом концепции. Однако непрерывное обучение создает серьёзные проблемы, связанные с обеспечением по-прежнему корректной работы ИИ-системы по мере её обучения. Необходимо проведение верификации находящейся в промышленной эксплуатации системы, а также сбор эксплуатационных данных для включения их в набор обучающих данных в том случае, если ИИ-система будет обновляться в какой-то момент времени в будущем.

Ввиду риска катастрофического забывания, использование непрерывного обучения подразумевает наличие способности учиться с течением времени, приспосабливаясь к новым наблюдениям, сделанным на основе текущих данных, но сохраняя в то же время предыдущие знания.

К особенностям непрерывного обучения относятся:

- обучение с течением времени в динамичной среде (в идеале, в открытом мире);
- расширение ранее полученных знаний за счет изучения новых знаний с целью повышения эффективности и производительности (либо с использованием новых данных, либо за счет рассуждений на основе существующих знаний);
- обнаружение новых аспектов задачи, которые необходимо изучить, и их постепенное изучение;
- обучение «на рабочем месте» т.е. обучение во время работы системы в режиме промышленной эксплуатации.

5.12 Примеры алгоритмов машинного обучения

5.12.1 Нейронные сети

5.12.1.1. Общие положения

Нейронные сети стремятся имитировать интеллектуальные способности наблюдения, обучения, анализа и принятия решений в отношении сложных проблем. Ввиду этого источником идей при проектировании нейронных сетей служит то, как нейроны соединены друг с другом в мозге людей и животных. Структура нейронных сетей состоит из взаимосвязанных обрабатывающих элементов, называемых нейронами. Каждый нейрон получает несколько входных значений и вырабатывает только одно выходное значение. Нейроны организованы в слои, при этом выходные данные одного слоя становятся входными данными для следующего слоя. Каждому соединению (связи) между нейронами назначается весовой коэффициент, отражающий важность соответствующего входного сигнала. Нейронная сеть «обучается», тренируясь на известных входных данных, сравнивая фактически полученный результат с ожидаемым и используя вычисленные ошибки для корректировки весовых коэффициентов. В результате вырабатывающие правильные ответы связи усиливаются, а те, что вырабатывают неправильные ответы, ослабевают.

В настоящем стандарте глубокое обучение определяется как подход к созданию многогранных иерархических представлений посредством обучения нейронных сетей с большим количеством скрытых слоев. Такой процесс позволяет нейронной сети постепенно уточнять конечный результат. Глубокое обучение может уменьшить или исключить необходимость в проектировании признаков (feature engineering), поскольку наиболее релевантные признаки выявляются автоматически.

Глубокое обучение может потребовать значительного времени вычислений и вычислительных ресурсов.

Существует множество «архитектур» нейронных сетей (являющихся, по сути, способами организации нейронов) - и это активная область исследований, в которой продолжает разрабатываться и внедряться ряд новых архитектур нейронных сетей. В числе примеров архитектур нейронных сетей можно назвать следующие:

- нейронная сеть прямого распространения (с прямой связью);
- рекуррентная нейронная сеть;
- свёрточная нейронная сеть.

Эти архитектуры нейронных сетей описаны в подразделах с 5.12.1.2 по 5.12.1.4.

Примечание – Дополнительную информацию о нейронных сетях см. в стандарте ИСО/МЭК 23053.

5.12.1.2. Нейронная сеть прямого распространения (с прямой связью)

Нейронная сеть прямого распространения (FFNN) - самая простая архитектура нейронной сети. В этом случае информация передаётся только в одном направлении, от входа к выходу. Отсутствуют связи между нейронами одного и того же слоя. Два соседних слоя обычно могут быть «полностью соединены» в том смысле, что каждый нейрон в одном слое соединён с каждым нейроном в следующем слое.

5.12.1.3. Рекуррентная нейронная сеть

5.12.1.3.1. Общие положения

Рекуррентная нейронная сеть [21] имеет дело с входными данными, которые поступают в виде упорядоченной последовательности, т.е.

порядок входных данных в последовательности имеет значение. Примерами таких входных данных могут служить динамические последовательности, такие как звуковые и видеопотоки, но также и статические последовательности, такие как текст или даже отдельные изображения. Рекуррентные нейронные сети имеют узлы, которые получают входную информацию с предыдущего уровня, но также учитывают собственную информацию с предыдущего прохода. Рекуррентные нейронные сети обладают свойством запоминания состояния, на которое влияет прошлое обучение. Рекуррентные нейронные сети широко используются для распознавания речи, машинного перевода, прогнозировании временных рядов и распознавании изображений. Дополнительную информацию о рекуррентных нейронных сетях см. в стандарте ИСО/МЭК 23053.

5.12.1.3.2. Сеть с архитектурой долгой краткосрочной памяти

Сеть с архитектурой долгой краткосрочной памяти (LSTM-сеть) - это тип рекуррентной нейронной сети, разработанный для задач, требующих запоминания информации как на более длинных, так и на более коротких интервалах времени, что делает их подходящими для изучения долговременных связей. Они были введены для решения проблемы «исчезающего» (затухающего) градиента в рекуррентных нейронных сетях, связанной с использованием алгоритма обратного распространения ошибок [22].

LSTM-сети могут обучаться сложным последовательностям, например, писать в стиле Шекспира или сочинять музыку. Дополнительную информацию о LSTM-сетях см в стандарте ИСО/МЭК 23053.

5.12.1.4. Свёрточная нейронная сеть

Свёрточная нейронная сеть – это нейронная сеть, которая включает по крайней мере один слой свёртки для фильтрации полезной информации из входных данных. В число типичных применений подобных сетей входят распознавание изображений, разметка видеоматериалов и обработка естественного языка. Дополнительную информацию о свёрточных нейронных сетях см. в стандарте ИСО/МЭК 23053.

5.12.2 Байесовские сети

Байесовские сети - это модели на основе графов, используемые для прогнозирования зависимостей между переменными. Их можно использовать для определения вероятностей причин или переменных, которые могут повлиять на результат. Подобная причинно-следственная связь очень полезна в таких приложениях, как медицинская диагностика. В числе других полезных приложений байесовских сетей можно назвать анализ данных, работу с неполными данными и смягчение последствий чрезмерной подгонки моделей к данным (перетренированности). Байесовские сети полагаются на байесовскую вероятность: вероятность события зависит от степени уверенности в этом событии. Дополнительную информацию о байесовских сетях можно найти в [20] и ИСО/МЭК 23053.

5.12.3 Деревья решений

Деревья решений используют древовидную структуру решений для кодирования возможных результатов. Алгоритмы деревьев решений широко используются в задачах классификации и регрессии. Дерево формируется из узлов решений (узлов принятия решений) и листовых вершин. Из каждого узла решения выходят как минимум две ветви, в то время, как листовые вершины представляют собой окончательное

решение или классификацию. Узлы, как правило, упорядочены, начиная с решений, наиболее сильно влияющих на результат. Для получения наилучшего результата входные данные должны отражать различные факторы. Деревья решений аналогичны блок-схемам, в которых в каждом узле принятия решения может быть задан вопрос для определения ветви, к которой следует перейти.

5.12.4 Метод опорных векторов

Метод (машина) опорных векторов (SVM) - это метод машинного обучения, широко используемый в задачах классификации и регрессии. В рамках SVM проводится разметка неделимых элементов данных, которые при этом разделяются на две категории; в дальнейшем метод относит новые неделимые элементы данных к той или иной категории. Алгоритмы SVM - это алгоритмы классификации «максимального расстояния». Они определяют гиперплоскость, разделяющую два класса, находящиеся выше и ниже неё, обеспечивая максимальное расстояние (зазор) между классифицирующей гиперплоскостью и ближайшими точками данных. Ближайшие к границе точки называются опорными векторами. В методе SVM расстояние по нормали от опорных векторов до гиперплоскости составляет половину зазора. Обучение SVM включает максимизацию зазора с учетом данных, принадлежащих к различным категориям, находящимся на противоположных сторонах от разделяющей гиперплоскости. SVM также используют ядерные функции для отображения данных из исходного пространства в пространство большей размерности (иногда бесконечномерное), в котором и будет выбрана классифицирующая гиперплоскость.

Такие SVM-методы с жестким зазором редко используются на практике. Классификатор с жестким зазором работает только в том случае, если данные линейно разделимы. Достаточно одному

неделимому элементу данных оказаться на неправильной стороне гиперплоскости, и задача построения классификатора не может быть решена.

Классификаторы с мягким зазором, напротив, допускают нарушение границы неделимыми элементами данных (т.е. те могут располагаться на неверной стороне от гиперплоскости). Классификаторы с мягким зазором стремятся обеспечить максимальный зазор при одновременном ограничении нарушений зазора.

Примерами применения SVM являются задачи категоризации неразмеченных данных, прогнозирования и распознавания образов. Цель SVM при использовании в задачах регрессионного анализа является обратной цели SVM-классификатора. В ходе регрессионного анализа цель SVM заключается в том, чтобы разместить как можно больше неделимых элементов данных внутри зазора, одновременно ограничивая нарушения зазора (т.е. появление неделимых элементов данных вне зазора).

5.13 Автономность, гетерономия и автоматизация

ИИ-системы можно сравнивать по уровню автоматизации и по наличию внешнего контроля и управления. На одном конце спектра находится полностью автономная система, на другом - система, полностью контролируемая и управляемая человеком, а между ними – различные степени гетерономии. В таблице 1 показана взаимосвязь между автономией, гетерономией и автоматизацией, включая «нулевой случай» отсутствия автоматизации.

Таблица 1 — Связь между автономией, гетерономией и автоматизацией

<table>
<tr><th></th><th></th><th>Степень автоматизации</th><th>Комментарии</th></tr>
<tr><td rowspan="3">Автоматизированная система</td><td>Автономная</td><td>6 - Автономия</td><td>Система способна модифицировать свою целевую область применения или свои цели без внешнего вмешательства, управления или надзора.</td></tr>
<tr><td rowspan="3">Гетерономная</td><td>5 - Полная автоматизация</td><td>Система способна полностью выполнять свою миссию без внешнего вмешательства</td></tr>
<tr><td>4 - Высокая степень автоматизации</td><td>Система выполняет часть своей миссии без внешнего вмешательства</td></tr>
<tr><td></td><td>3 - Условная автоматизация</td><td>Система обеспечивает стабильные и соответствующие требованиям показатели эффективности и производительности, при этом внешний агент готов при необходимости взять управление на себя</td></tr>
</table>

Окончание таблицы 1

		Степень автоматизации	Комментарии
		2 - Частичная автоматизация	Некоторые функции системы полностью автоматизированы, в то время как система в целом остаётся под контролем и управлением внешнего агента.
		1 - Помощь	Система помогает оператору
		0 - Автоматизация отсутствует	Оператор полностью контролирует систему и управляет ею

Примечание – В юриспруденции под автономией понимается способность к самоуправлению. Использование термина «автономный» в таком смысле, однако, является некорректным применительно к автоматизированным ИИ-системам, поскольку даже самые продвинутые ИИ-системы не являются полностью самоуправляемыми. Скорее можно сказать, что ИИ-системы работают на основе алгоритмов, и в остальном подчиняются командам операторов. По этим причинам в настоящем стандарте популярный термин «автономный» не используется для описания автоматизации [30].

В число критериев, применимых для классификации систем по данной шкале, входят следующие:

- наличие или отсутствие внешнего надзора, осуществляемого либо оператором-человеком («человек в контуре управления»), либо иной автоматизированной системой;

- имеющаяся у системы степень понимания ситуации, включая полноту и способность применять на практике (операционализировать) имеющуюся у системы модель состояний её окружения; а также

уверенность, с которой система может рассуждать и действовать в своём окружении;

- степень способности реагировать и быстроты реагирования, включая, в том числе, способность системы заметить изменения в своём окружении, отреагировать на них, а также способность спрогнозировать будущие изменения;

- будет ли система продолжать функционировать вплоть до выполнения (или же продолжит работу и после этого) конкретной задачи или наступления конкретного события в её окружении (примером могут служить задача или событие, имеющие отношение к достижению цели; превышение лимитов по времени и другие механизмы);

- степень адаптивности к внутренним или внешним изменениям, потребностям и движущим силам;

- способность оценивать свои собственные показатели производительности и/или пригодность, включая оценку посредством сопоставления с предварительно поставленными целями;

- способность принимать решения и планировать «на упреждение», с учётом целей системы, мотивации и движущих сил.

В некоторых случаях машина вместо замены труда человека будет его функционально дополнять – это называется объединением человека и машины в команду. Такое может произойти как в качестве побочного эффекта развития ИИ, так и в результате целенаправленной разработки ИИ-системы с целью создания человеко-машинной команды. Системы, направленные на то, чтобы дополнять и расширять когнитивные возможности человека, иногда называют системами «усиления интеллекта» (intelligence augmentation).

В целом, наличие подотчётного надзора во время функционирования ИИ-системы может помочь с обеспечением её работы

таким образом, как это предполагалось, и с предотвращением нежелательных воздействий на заинтересованные стороны.

5.14 Интернет вещей и киберфизические системы

5.14.1 Общие положения

Искусственный интеллект всё чаще используется в качестве одного из компонентов во встроенных системах, таких как системы интернета вещей и киберфизические системы - либо для анализа поступающих с датчиков (сенсоров) потоков информации о физическом мире, либо для подготовки прогнозов и принятия решений в отношении физических процессов, на основе которых на исполнительные устройства (приводы) подаются соответствующие команды с целью управления этими физическими процессами или оказания на них влияния.

5.14.2 Интернет вещей

Интернет вещей (IoT) – это инфраструктура взаимосвязанных объектов, людей, систем и информационных ресурсов вместе с сервисами, которые обрабатывают и реагируют на информацию, поступающую из материального и виртуального миров (см. п. 3.1.18). По своей сути IoT-система представляет собой сеть узлов, оснащённых как датчиками, которые измеряют свойства физических объектов, а затем передают данные, относящиеся к этим измерениям; так и приводами, которые изменяют свойства физических объектов в ответ на цифровые входящие сигналы.

Примерами IoT-систем могут служить системы медицинского мониторинга и системы мониторинга состояния атмосферы. Здесь результатом работы систем является информация, предназначенная для оказания помощи людям в их деятельности (например, используемая для

предупреждения медицинского персонала или подготовки для людей прогнозов погоды).

IoT-системы включают взаимосвязанные ИТ-системы, взаимодействующие с физическими объектами. Фундаментальную роль при построении IoT-систем играют цифровые IoT-устройства в виде датчиков и взаимодействующих с физическими объектами исполнительных устройств (приводов). Датчик измеряет один или несколько параметров одного или нескольких физических объектов, и выдаёт данные измерений, которые могут быть переданы по сети. Привод изменяет одно или несколько свойств физического объекта в ответ на полученные по сети корректные входные данные. Как датчики, так и приводы могут быть различных типов, например, термометры, акселерометры, видеокамеры, микрофоны, реле, обогреватели, роботы и промышленное оборудование для производства или управления процессами. Для получения дополнительной информации см. стандарт ИСО/МЭК 30141.

Искусственный интеллект может сыграть важную роль в контексте IoT-систем. Сюда входит анализ входящих данных и принятие решений, которые могут помочь в достижении целей системы, таких как управление физическими объектами и физическими процессами, посредством предоставления в реальном времени контекстуализированной прогнозной информации.

5.14.3 Киберфизические системы

Киберфизические системы (КФС) – это системы, аналогичные IoT-системам, но такие, в которых логика управления применяется к поступающим с датчиков входным данным с целью направлять работу исполнительных устройств и тем самым влиять на процессы, происходящие в физическом мире.

Примером киберфизической системы может служить робот. В этом случае поступающие от датчиков входные данные напрямую используются для управления роботом и для выполнения действий в физическом мире.

Робототехника охватывает все виды деятельности, связанные с проектированием, сборкой, производством, управлением и использованием роботов для выполнения различных прикладных задач. Робот состоит из электронных, механических, программных компонентов и встроенных программ, тесно взаимодействующих друг с другом ради достижения целей, поставленных в рамках конкретной прикладной задачи. Роботы обычно содержат датчики для оценки их текущей ситуации; блоки обработки для обеспечения контроля и управления посредством анализа и планирования действий; приводы для выполнения действий. Промышленные роботы, устанавливаемые на производственных участках, запрограммированы на то, чтобы в точности раз за разом и без каких-либо отклонений повторять одни и те же траектории и действия. Сервисные роботы и роботы для совместной работы с человеком (коллаборативные роботы, коботы) должны адаптироваться к изменяющимся ситуациям и динамическим средам. Программирование этой гибкости чрезвычайно сложно из-за всей той изменчивости, с которой приходится иметь дело.

Как компоненты киберфизических систем, ИИ-системы могут быть частью управляющего программного обеспечения и процесса планирования в рамках парадигмы «измеряй, планируй, действуй», обеспечивая роботам возможность приспосабливаться к ситуации в случае появления препятствий или в случае перемещения целевых объектов. Объединение робототехники с ИИ-системами в качестве компонентов делает возможным автоматическое физическое взаимодействие с объектами, окружающей средой и людьми.

5.15 Доверие к ИИ-системам

5.15.1 Общие положения

Под способностью ИИ-систем вызывать доверие понимаются качества и свойства, которые помогают соответствующим заинтересованным сторонам понять, соответствует ли ИИ-система их ожиданиям. Эти качества и свойства могут помочь заинтересованным сторонам убедиться в том, что:

- ИИ-системы были должным образом спроектированы, и было подтверждено их соответствие действующим правила и стандартам. Это подразумевает обеспечение уверенности в качестве и робастности;

- ИИ-системы созданы в интересах соответствующих заинтересованных сторон, имеющих согласованные цели. Это подразумевает осведомленность заинтересованных сторон о механизмах работы ИИ-алгоритмов и понимание ими общих принципов функционирования ИИ-системы. Это также подразумевает обеспечение уверенности в квалификации и/или проведение сертификации процессов разработки и эксплуатации ИИ в соответствии с нормативно-правовыми требованиями и отраслевыми стандартами, когда таковые имеются;

- надлежащим образом идентифицированы ответственные и подотчётные за ИИ-системы стороны;

- ИИ-системы разрабатываются и эксплуатируются с учётом соответствующих региональных интересов.

Дополнительную информацию см. в техническом отчёте ИСО/МЭК ТО 24028.

5.15.2 Робастность ИИ-систем

Применительно к ИИ-системам, под робастностью понимается их способность при любых обстоятельствах поддерживать предполагавшийся их разработчиками уровень показателей производительности. Примером робастности является способность системы выполнять свои функции в приемлемых пределах, несмотря на внешнее вмешательство или жёсткие условия окружающей среды. Робастность может охватывать другие свойства, такие как жизнеспособность и надёжность. Надлежащее функционирование ИИ-системы непосредственно связано либо приводит к безопасности заинтересованных в ней сторон в данной среде или контексте (см. ИСО/МЭК ТО 24028).

Например, робастная ИИ-система на основе машинного обучения может быть способна выполнять обобщение для зашумленных входных данных, предотвращая, например, чрезмерную подгонку модели к данным (перетренированность). Одним из вариантов обеспечения робастности является обучение модели (моделей) с использованием больших наборов обучающих данных, включающих зашумленные обучающие данные (см. ИСО/МЭК ТО 24028).

Наличие свойства робастности говорит о способности (или неспособности) системы обеспечивать сопоставимые показатели производительности на нетипичных данных - в отличие от данных, ожидаемых в ходе типичных операций; или на входных данных, непохожих на те, на которых система была обучена (см. ИСО/МЭК ТО 24029-1).

При обработке входных данных от ИИ-системы ожидается, что она будет генерировать прогнозы (являющиеся её результатами) в рамках некоторого приемлемого, согласованного и/или эффективного диапазона. Даже если эти результаты не являются идеальными, система всё ещё

может считаться робастной. ИИ-систему, у которой результаты обработки входных данных не укладываются в этот приемлемый, согласованный и/или эффективный диапазон, робастной считать нельзя.

Робастность может по-разному интерпретироваться для различных типов ИИ-систем, например:

- робастность регрессионной модели – это способность иметь приемлемые метрики амплитуды отклика при любом корректном входном значении;

- робастность классификационной модели – это способность избегать появления новых ошибок классификации при переходе от типичных входных значений к входных значениям, находящимся в определенном диапазоне, отличающемся от типичных значений.

5.15.3 Надёжность ИИ-систем

Надёжность – это способность системы или объекта в этой системе выполнять требующиеся от него функции в заданных условиях в течение установленного периода времени (см. ИСО/МЭК 27040).

Под надёжностью ИИ-системы понимается её способность последовательно и корректно выдавать на стадии эксплуатации (см. п.6.2.6) требуемые прогнозы (см. п.3.1.27), рекомендации и решения.

На надёжность могут повлиять, как минимум, робастность, способность обобщать, последовательность и жизнеспособность ИИ-системы. Предполагается, что все входные данные и настройки окружения, соответствующие установленным критериям, должны правильно обрабатываться в ходе функционирования ИИ-системы. Некоторые из входных данных могут отличаться от тех данных, что использовались на стадии разработки, но их появление возможно в ходе эксплуатации системы. Резервирование ИИ-системы или её компонентов также повышает надежность, обеспечивая реализации логики деловых

процессов, которые ведут себя так же, как и исходная реализация. Резервная система включается в работу в случае сбоя ИИ-системы.

Надёжность может способствовать функциональной безопасности ИИ-системы в том смысле, что в соответствии с требованиями заинтересованных сторон для защиты системы или её части от определенного вида отказа нужны автоматические меры защиты и операции.

5.15.4 Жизнеспособность ИИ-систем

Жизнеспособность – способность системы быстро восстанавливать рабочее состояние после инцидента. Отказоустойчивость – это способность системы продолжать функционировать (возможно, с пониженными возможностями) при возникновении в системе сбоев, отказов и неисправностей.

В случае ИИ-систем, как и в случае других типов информационных систем, сбои и отказы оборудования могут повлиять на правильное выполнение алгоритма.

Надёжность и жизнеспособность взаимосвязаны, однако ожидаемые уровни обслуживания и ожидания различны, причём установленные заинтересованными сторонами ожидания в отношении жизнеспособности, возможно, ниже. В случае определённых типов сбоев и отказов жизнеспособная система может предложить пониженный уровень функционирования, который может быть приемлемым для заинтересованных сторон. Жизнеспособные системы также должны, по мере необходимости, предусматривать способы восстановления работоспособности.

5.15.5 Управляемость ИИ-систем

Управляемость – это свойство ИИ-системы, означающее возможность человека или иного внешнего агента вмешиваться в функционирование системы. Управляемость может достигаться посредством предоставления надёжных механизмов, с помощью которых агент может взять на себя управление ИИ-системой.

Ключевым аспектом управляемости является определение того, какие агенты (например, поставщики продуктов или сервисов, поставщик компонентов ИИ, пользователь, и/или регулирующий орган) и какими компонентами ИИ-системы могут управлять.

Дополнительную информацию об управляемости можно найти в техническом отчёте ИСО/МЭК ТО 24028:2020, п.9.4.

5.15.6 Объяснимость ИИ-систем

Объяснимость - свойство ИИ-системы предоставлять информацию о влияющих на её результаты существенных факторах в понятном для людей виде. Объяснимость может быть особенно важна, когда принимаемые ИИ-системой решения затрагивают интересы физических лиц. Люди склонны не доверять решению, если не могут понять, что к нему привело, особенно если это решение каким-либо образом неблагоприятно для них лично (например, отказано в предоставлении кредита).

Объяснимость также может быть полезным инструментом при валидации ИИ-системы, даже если принимаемые решения напрямую на людей не влияют. Например, если ИИ-система анализирует представленную на изображении сцену с целью идентификации в ней объектов, то может быть полезно увидеть объяснение причин решения, касающегося содержания сцены - как способ убедиться в том, что результаты идентификации действительно соответствуют тому, что

утверждается. В истории ИИ-систем известны примеры, когда, при отсутствии такого рода объяснений, впоследствии оказалось, что ИИ-система идентифицировала некоторые объекты в сцене на основе присутствовавших в обучающих данных случайных корреляций.

Объяснимость может быть проще обеспечить для одних типов ИИ-систем, чем для других. Так, объяснимость глубоких нейронных сетей может представлять проблему, поскольку сложность системы может затруднить предоставление осмысленного объяснения того, как система пришла к решению.

Основанные на правилах алгоритмы, такие как символьные методы и деревья решений, часто считаются хорошо объяснимыми, поскольку сами правила напрямую позволяют дать определённые объяснения. Тем не менее, эти объяснения могут стать менее понятными по мере того, как растут размеры и сложность таких моделей.

5.15.7 Предсказуемость ИИ-систем

Предсказуемость – это свойство ИИ-системы, дающее возможность заинтересованным сторонам делать надёжные предположения о результатах её работы. Предсказуемость играет важную роль в обеспечении приемлемости ИИ-систем и часто упоминается в дебатах по этике в отношении ИИ-систем. Доверие к технологиям часто основано на предсказуемости: системе доверяют, если её пользователи способны предсказать, как система поведёт себя в определенной ситуации, - пусть даже эти пользователи и не смогут объяснить факторы, лежащие в основе поведения системы. Напротив, пользователи могут потерять доверие к системе, если та начнёт работать неожиданным образом в ситуациях, когда правильный ответ кажется очевидным.

Тем не менее, имеется несколько проблем с наивным представлением о предсказуемости, основанном на идее о том, что

человек должен иметь возможность предсказывать поведение ИИ-системы:

- определение, непосредственно опирающееся на понимание человеком, по своей сути субъективно. Определение предсказуемости должно использовать объективные, количественно измеримые критерии.

- должна быть возможность установить доверие к ИИ-системе в случае, когда, например, один-единственный человек не может предсказать её точное поведение во всех ситуациях. Статистическая гарантия уместности поведения ИИ-системы может быть более полезной. Обоснованием такого утверждения служит то, что многие методы машинного обучения выдают неизбежно непредсказуемые результаты.

Предсказуемость взаимосвязана с точностью. Повышающие точность методы могут снизить вероятность того, что ИИ-системы будут выдавать непредсказуемые результаты.

5.15.8 Прозрачность ИИ-систем

Прозрачность ИИ-систем поддерживает человеко-ориентированные цели системы и является темой продолжающихся исследований и дискуссий. Обеспечение прозрачности в отношении ИИ-системы может включать передачу заинтересованным сторонам соответствующей информации о системе (например, сведений о целях, известных ограничениях, определениях, проектных решениях, предположениях, характеристиках, моделях, алгоритмах, методах обучения и процессах обеспечения уверенности в качестве). Кроме того, обеспечение прозрачности ИИ-системы может включать информирование заинтересованных сторон о подробностях, касающихся данных, используемых при создании системы (например, о том, какие, где, когда и почему собираются данные, и как они используются), а также защиты персональных данных - вместе со сведениями о назначении системы и о

том, как она была построена и развёрнута. Обеспечение прозрачности также может включать информирование заинтересованных сторон об обработке, проводимой с целью принятия соответствующих решений, и об используемом при этом уровне автоматизации.

Примечание – Раскрытие определённой информации в интересах обеспечения прозрачности может противоречить требованиям по обеспечению безопасности, конфиденциальности и защите неприкосновенности частной жизни (персональных данных).

5.15.9 Предвзятость и справедливость ИИ-систем

Понятие «неодинаковый подход» (bias), в зависимости от контекста, может подразумевать как «предвзятость, необъективность», так и просто «дифференцированный подход».

В области искусственного интеллекта под «дифференцированным подходом» понимается идея о том, что разные ситуации требуют различного подхода. В этом смысле дифференцированный подход позволяет системам машинного обучения судить о том, что одна ситуация отличается от другой, и вести себя по-разному. Таким образом, дифференцированный подход является фундаментальным фактором для процесса машинного обучения и для адаптации поведения к конкретной ситуации, с которой приходится сталкиваться

Однако в социальном контексте под термином «предвзятость» часто понимается представление о том, что определенные различия в отношении являются несправедливыми. Во избежание путаницы в контексте ИИ вместо данного термина используется термин «несправедливость» (unfairness), означающий неоправданную дифференциацию в отношении и обработке, отдающую предпочтение определённым группам, в результате чего те выигрывают по сравнению с другими группами. Несправедливое поведение ИИ-системы может

привести к неуважению к установленным фактам и к сложившимся убеждениям и нормам, следствием чего могут стать фаворитизм и/или дискриминация.

Несмотря на то, что определенная дифференцированность в отношении необходима для надлежащей работы ИИ-системы, в ИИ-систему может быть непреднамеренно введена нежелательная предвзятость, что может привести к несправедливым результатам работы системы. Источники нежелательной предвзятости в ИИ-системах взаимосвязаны и включают в себя когнитивную предвзятость человека, предвзятость в данных и предвзятость, которую вводят инженерные решения. Предвзятость в обучающих данных является основным источником предвзятости в ИИ-системах. Когнитивная предвзятость человека может повлиять на решения, касающиеся сбора и обработки данных, архитектуры системы, обучения моделей, а также на другие решения, принимаемые в ходе разработки.

Минимизация нежелательной предвзятости в ИИ-системах является сложной задачей, однако выявление и устранение предвзятости возможны, см. [13].

5.16 Верификация и валидация ИИ-систем

Верификация является подтверждением того, что система была построена корректно и выполняет установленные требования. Валидация является подтверждением, посредством представления объективных свидетельств того, что были выполнены требования в отношении конкретного предполагаемого использования или применения. Относящиеся к верификации и валидации соображения включают следующее:

- некоторые системы являются полностью верифицируемыми (т.е. могут быть верифицированы как все компоненты системы по отдельности, как и система в целом).

- некоторые системы являются частично верифицируемыми и частично способными пройти валидацию (если, например, по крайней мере один компонент системы может быть индивидуально верифицирован, а остальные компоненты и система в целом способны пройти валидацию).

- некоторые системы являются неверифицируемыми, но способными пройти валидацию (если, например, ни один компонент системы не может быть верифицирован, однако, как все компоненты системы, так и система в целом способны пройти валидацию).

- некоторые системы являются неверифицируемыми и способными лишь частично пройти валидацию (если, например, ни один компонент системы не может быть верифицирован, но хотя бы один из компонентов способен индивидуально пройти валидацию).

- некоторые системы являются неверифицируемыми и не способными пройти валидацию (если, например, ни один компонент системы не способен пройти ни верификацию, ни валидацию).

5.17 Использование ИИ-систем в нескольких юрисдикциях

ИИ-системы могут развертываться и эксплуатироваться в юрисдикциях, отличных от тех, в которых система была спроектирована и/или изготовлена. Разработчики и производители ИИ-систем должны понимать, что применяемые нормативные правовые требования в разных юрисдикциях могут различаться.

Например, от автомобиля, произведённого в одной юрисдикции, может потребоваться соответствие отличающимся нормативным

правовым требованиям, чтобы было разрешено ввезти его на территорию с другой юрисдикцией.

Кроме того, ИИ-системы обычно требуют сбора, обработки и использования данных на стадиях разработки и эксплуатации ИИ-системы, а также уничтожения данных на стадии вывода из эксплуатации. Разработчики, производители и пользователи ИИ-систем должны знать, что нормативные правовые требования в отношении сбора, использования и уничтожения данных также могут различаться в разных юрисдикциях.

Чтобы смягчить последствия неодинаковости нормативных правовых требований, разработчики и производители ИИ-систем могут использовать одну или несколько из следующих мер:

- выявите применимые нормативные правовые требования, под которые ИИ-система может подпадать на этапе подготовки. В их число должны быть включены нормативные правовые требования, касающиеся сбора, использования и уничтожения данных;

- разработайте план исполнения применимых нормативных правовых требований той юрисдикции (юрисдикций), в которой предполагается развернуть и эксплуатировать ИИ-систему;

- разработайте план мониторинга исполнения нормативных правовых требований во время проектирования и разработки, развертывания, эксплуатации и вывода из эксплуатации ИИ-системы.

- разработайте план мониторинга любых изменений в нормативных правовых требованиях и реагирования на них.

- внедрите гибкие подходы к проектированию, развертыванию и эксплуатации.

5.18 Социальное воздействие

ИИ-системы несут с собой ряд рисков, категории которых определяется тяжестью потенциальных последствий отказов, сбоев и неожиданного поведения. В число существенных факторов для оценки уровня риска входят следующие:

- тип пространства действий, в рамках которого система функционирует (например, это могут быть рекомендации или же прямые действия, выполняемые системой в её окружении);
- присутствие или отсутствие внешнего надзора;
- тип внешнего надзора (автоматизированный или ручной);
- этическая значимость задачи или области применения;
- уровень прозрачности решений или этапов обработки;
- степень автоматизации системы.

Например, применяемая в не имеющей этической значимости области ИИ-система, которая лишь даёт рекомендации и не может действовать самостоятельно, может быть отнесена к категории низкого риска. И наоборот, ИИ-система может сбыть отнесена к категории высокого риска, если её действия оказывают прямое воздействие на жизни людей, если она действует без внешнего надзора, а её процесс принятия решений является непрозрачным.

Примечание - В конкретных областях применения ИИ-систем могут быть применимы дополнительные нормативно-правовые требования, политики и стандарты, которые могут выходить за рамки описанного в данном разделе анализа воздействия.

5.19 Роли заинтересованных сторон

5.19.1 Общие положения

Как показано на рисунке 2, в области искусственного интеллекта заинтересованные стороны могут выполнять ряд ролей и суб-ролей. Эти роли и суб-роли описаны в подразделах с 5.19.2 по 5.19.7.

Примечание – Организация или субъект могут взять на себя выполнение более одной роли или суб-роли.

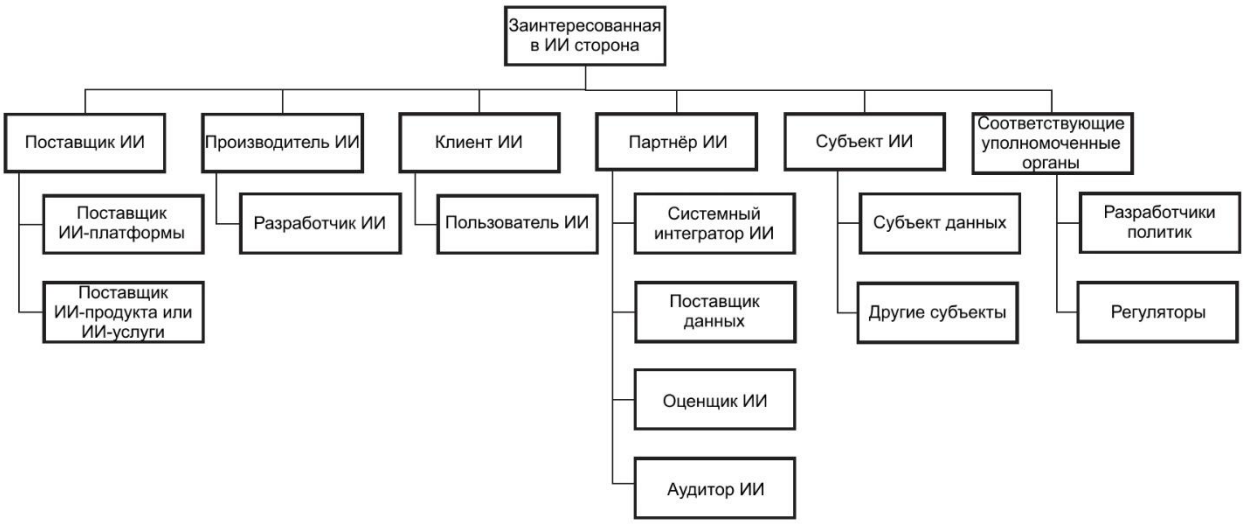

Рисунок 2 — Роли и суб-роли заинтересованных в ИИ сторон

5.19.2 Поставщик ИИ

5.19.2.1. Общие положения

Поставщик ИИ – это организация или субъект, который предоставляет продукты и/или услуги, использующие одну или несколько ИИ-систем. В число поставщиков ИИ входят поставщики ИИ-платформ и поставщики ИИ-продуктов или ИИ-услуг.

5.19.2.2. Поставщик ИИ-платформы

Поставщик ИИ-платформы – это организация или субъект, который предоставляет услуги (сервисы), позволяющие другим заинтересованным сторонам производить ИИ-продукты или ИИ-услуги.

5.19.2.3. Поставщик ИИ-продукта или ИИ-услуги

Поставщик ИИ-продукта или ИИ-услуги (сервиса) – это организация или субъект, который предоставляет ИИ-продукты или ИИ-услуги (сервисы), являющиеся либо непосредственно пригодными для использования клиентом ИИ или пользователем ИИ, либо предназначенные для интеграции в систему, использующую ИИ-компоненты наряду с компонентами без ИИ.

5.19.3 Производитель ИИ

5.19.3.1. Общие положения

Производитель ИИ – это организация или субъект, который проектирует, разрабатывает, тестирует и развёртывает продукты или предоставляет услуги, использующие одну или несколько ИИ-систем.

5.19.3.2. Разработчик ИИ

Разработчик ИИ – это организация или субъект, который занимается разработкой ИИ-услуг и ИИ-продуктов. Примерами разработчиков ИИ могут служить (не ограничиваясь ими):

- проектировщик модели: субъект, который получает данные и постановку задачи и создаёт ИИ-модель;

- имплементатор модели: субъект, который получает ИИ-модель и указывает, какие вычисления следует выполнять (давая указания, как использовать модель и на каких вычислительных ресурсах – например, на процессорах типа CPU, GPU, ASIC, FPGA);

- верификатор вычислений: субъект, который проверяет, что вычисления выполняются в соответствии с проектным решением;

- верификатор модели: субъект, который проверяет, что показатели производительности ИИ-модели соответствуют проектному решению.

5.19.4 Клиент ИИ

5.19.4.1. Общие положения

Клиент ИИ – это организация или субъект, который использует ИИ-продукт или ИИ-услугу непосредственно, либо путём их предоставления пользователям ИИ.

5.19.4.2. Пользователи ИИ

Пользователь ИИ – это организация или субъект, который использует ИИ-продукты или ИИ-услуги.

5.19.5 Партнёр ИИ

5.19.5.1. Общие положения

Партнёр ИИ – это организация или субъект, который предоставляет услуги в сфере искусственного интеллекта. Партнеры ИИ могут выполнять техническую разработку ИИ-продуктов или ИИ-услуг, проводить их тестирование и валидацию, проводить аудит применения ИИ, оценивать ИИ-продукты или ИИ-услуги, а также выполнять другие задачи. Примеры различных видов ИИ-партнёров обсуждаются в последующих подразделах.

5.19.5.2. Системный интегратор ИИ

Системный интегратор ИИ – это организация или субъект, который занимается интеграцией ИИ-компонентов в более крупные системы, потенциально также включающие компоненты без ИИ.

5.19.5.3. Поставщик данных

Поставщик данных – это организация или субъект, который предоставляет данные, используемые ИИ-продуктами или ИИ-услугами.

5.19.5.4. Аудитор ИИ

Аудитор ИИ – это организация или субъект, занимающийся аудитом организаций, которые производят, предоставляют или используют ИИ-системы, с целью оценки соответствия стандартам, политикам и/или нормативным правовым требованиям.

5.19.5.5. Оценщик ИИ

Оценщик ИИ – это организация или субъект, который оценивает показатели производительности одной или нескольких ИИ-систем.

5.19.6 Субъект ИИ

5.19.6.1. Общие положения

Субъект ИИ – это организация или субъект, на который оказывают воздействие ИИ-система, ИИ-услуга или ИИ-продукт.

5.19.6.2. Субъект данных

Субъект данных – это организация или субъект, на который ИИ-системы оказывают следующее воздействие:

- субъект обучающих данных: В случае, когда относящиеся к организации или человеку данные используются для обучения ИИ-системы, возможны негативные последствия для безопасности и неприкосновенности частной жизни (защиты персональных данных) – последнее особенно актуально в том случае, когда субъект данных является физическим лицом.

5.19.6.3. Другие субъекты

Другими организациями или субъектами, на которых оказывают воздействие ИИ-система, ИИ-услуга или ИИ-продукт, могут быть, например, физические лица или сообщества. Примерами могут служить

потребители, которые взаимодействуют с социальной сетью, предоставляющей рекомендации с использованием ИИ; или же водители транспортных средств, оснащённых средствами автоматизации на основе ИИ.

5.19.7 Соответствующие уполномоченные органы

5.19.7.1. Общие положения

Соответствующими уполномоченными органами являются организации или субъекты, которые могут оказать влияние на ИИ-системы, ИИ-услуги или ИИ-продукты.

5.19.7.2. Разработчики политик

Разработчики политик (устанавливающие политики лица или органы) – это организации или субъекты, обладающие полномочиями устанавливать на международном, региональном, национальном или отраслевом уровне политики, способные оказать влияние на ИИ-системы, ИИ-услуги или ИИ-продукты.

5.19.7.3. Регуляторы

Регуляторы – это организации или субъекты, обладающие полномочиями устанавливать, реализовать и обеспечивать соблюдение нормативных правовых требований, в соответствии с намерениями политик, установленных разработчиками политик (5.17.9.2).

6 Жизненный цикл ИИ-системы

6.1 Модель жизненного цикла ИИ-системы

Модель жизненного цикла ИИ-системы описывает эволюцию ИИ-системы от возникновения замысла до вывода из эксплуатации. Данный

стандарт не предписывает какой-либо конкретной модели жизненного цикла. Вместо этого в нём основное внимание обращается на характерные для ИИ-систем процессы, которые могут происходить в течение жизненного цикла системы. Характерные для ИИ процессы и их хронологические последовательности могут иметь место на одной или нескольких стадиях жизненного цикла, а отдельные стадии жизненного цикла могут повторяться в течение жизненного цикла системы. Например, возможно неоднократное принятие решений о повторном прохождении стадий «проектирование и разработка» и «развёртывание» для разработки и внедрения исправлений ошибок и обновлений системы.

Модель жизненного цикла системы помогает заинтересованным сторонам создавать ИИ-системы более эффективно и продуктивно. При разработке модели жизненного цикла полезны международные стандарты, в том числе стандарты ИСО/МЭК/ИИЭР 15288 для систем в целом, ИСО/МЭК/ИИЭР 12207 - для программного обеспечения и ИСО/МЭК/ИИЭР 15289 - для документации на систему. Эти международные стандарты описывают процессы жизненного цикла для любых систем и не являются специфическими для ИИ-систем. На рисунке 3 показан пример стадий и высокоуровневых процессов, которые могут использоваться в жизненном цикле ИИ-систем. Стадии и процессы могут выполняться итеративно, что часто требуется в ходе разработки и эксплуатации ИИ-систем. Имеется ряд аспектов, которые следует принять во внимание при разработке модели жизненного цикла. Примерами таких аспектов являются:

- последствия для стратегического управления, возникающие вследствие разработки и/или использования ИИ-систем;

- последствия для безопасности и неприкосновенности частной жизни (защиты персональных данных) вследствие использования

больших объёмов данных, некоторые из которых могут быть «чувствительными» по своему характеру;

- угрозы безопасности, возникающие вследствие зависимого от данных процесса разработки системы;

- факторы прозрачности и объяснимости, включая наличие сведений о происхождении данных и способность дать объяснение того, как определяются результаты работы ИИ-системы.

На рисунке 3 приведен пример стадий и высокоуровневых процессов в модели жизненного цикла ИИ-системы. В Приложении А показано, как эта модель жизненного цикла ИИ-системы соотносится с определением жизненного цикла ИИ-системы, разработанным Организацией экономического сотрудничества и развития (ОЭСР).

Рисунок 3 — Пример стадий и высокоуровневых процессов в модели жизненного цикла ИИ-системы

ИИ-системы отличаются от других типов информационных систем, что может повлиять на процессы в модели жизненного цикла. Например:

- большинство систем запрограммировано на то, чтобы вести себя точно определенным образом, обусловленным требованиями к ним и их спецификациями. ИИ-системы использующие машинное обучение, применяют методы обучения и оптимизации на основе данных для обработки сильно варьирующихся входных данных;

- традиционные программные приложения, как правило, предсказуемы, в то время, как предсказуемость ИИ-систем встречается не так часто;

- традиционные программные приложения также обычно являются верифицируемыми, в то время как оценка производительности ИИ-систем часто требует применения статистических подходов, и их верификация может быть проблематичной;

- ИИ-системы обычно нуждаются в ряде улучшающих итераций для достижения приемлемого уровня производительности.

Ключевым аспектом ИИ-систем является управление данными (охватывающее процессы и инструменты для комплектования данных, их аннотирования, подготовки, проверки качества, формирования выборки и аугментации).

Процессы разработки и тестирования для ИИ-систем также отличаются, поскольку эти процессы опять же опираются на данные. Всё становится ещё более сложным в случае ИИ-систем, использующих непрерывное обучение (также известное как продолжающееся обучение или обучение на протяжении всего жизненного цикла), где система обучается и на стадии эксплуатации и где требуется регулярное проведение тестирования.

Процесс управления версиями у ИИ-систем отличается от того, что применяется в случае традиционного программного обеспечения. Если в

случае традиционных программных приложений решается задача управления версиями кода и используются функции, определяющими различия между этими версиями, то в случае ИИ-систем различия между версиями включает в себя различия как в коде, так и в модели - а также различия в обучающих данных, если используется машинное обучение.

Некоторые процессы жизненного цикла ИИ-систем, которые отличаются от процессов жизненного цикла традиционного программного обеспечения, обсуждаются в п.6.2.

На рисунке 4 показан пример модели жизненного цикла для ИИ-системы. Возможны различные модели жизненного цикла в зависимости от различных методов разработки. На рисунке 4 приведена последовательность стадий жизненного цикла, для каждой из которых указаны процессы, являющиеся значимыми для ИИ-систем и требующие дополнительного анализа помимо того, что необходим в ходе разработки типичных систем без ИИ.

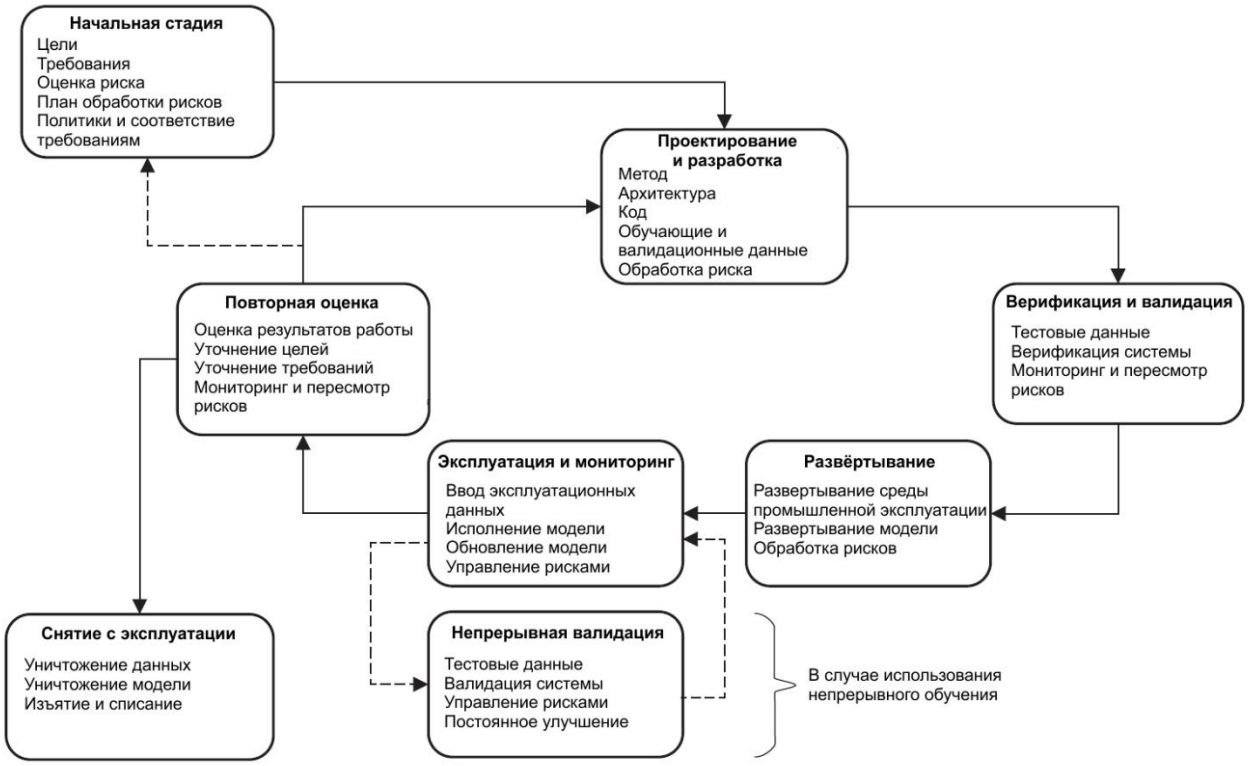

Рисунок 4 — Пример модели жизненного цикла для ИИ-системы с указанием специфических для ИИ процессов

Как показано на рисунке 4, разработка и эксплуатация ИИ-систем, как правило, носит более итеративный характер, чем в случае систем без ИИ. ИИ-системы склонны быть менее предсказуемыми, и обычно для достижения ИИ-системой своих целей требуются определенный опыт работы с ней и её настройка.

6.2 Стадии и процессы жизненного цикла ИИ-системы

6.2.1 Общие положения

Описанные ниже в составе каждой стадии процессы представляют собой репрезентативные примеры, поскольку конкретные процессы будут зависеть от конкретной ИИ-системы. Процессы могут выполняться в различном порядке, а в некоторых случаях - параллельно.

Данные процессы сами по себе не обязательно являются специфическими для ИИ, однако связанные с ИИ риски и возможности придают им в этом контексте особое значение.

6.2.2 Начальная стадия

Начальная стадия выполняется тогда, когда одна или несколько заинтересованных сторон решают превратить идею в реальную систему. Начальная стадия может включать несколько процессов и решений, которые приводят к решению перейти к стадии проектирования и разработки. В ходе жизненного цикла к начальной стадии возможно повторно вернуться в том случае, если новая информация будет выявлена на более поздних стадиях - например, может оказаться, что

система технически или финансово нереализуема. Примерами процессов, которые могут происходить на начальной стадии, являются:

Цели: Заинтересованные стороны должны определить, зачем необходимо разработать ИИ-систему. Какую проблему система решает? Какую потребность клиента система удовлетворяет, какие деловые возможности обеспечивает? Каковы метрики успешности?

Требования: Заинтересованные стороны должны сформировать набор требований к ИИ-системе, охватывающих весь её жизненный цикл. Неспособность подготовить требования к стадиям развертывания, эксплуатации и снятия с эксплуатации может привести к проблемам в будущем. Выявить потенциальные риски и непреднамеренные последствия создания и эксплуатации системы может помочь подход, предусматривающий привлечение ряда заинтересованных сторон и экспертов в различных предметных областях. Заинтересованные стороны должны позаботиться о том, чтобы требования к ИИ-системе обеспечивали достижение целей создаваемой ИИ-системы. Требования должны учитывать то, что многие ИИ-системы не являются предсказуемыми, а также воздействие, которое такая непредсказуемость может оказать на достижение целей. Заинтересованные стороны должны принять во внимание фактор нормативных правовых требований и обеспечить исполнение соответствующих обязательных политик при разработке и эксплуатации ИИ-систем.

Управление рисками: Организации должны оценивать связанные с ИИ риски на протяжении всего жизненного цикла ИИ-системы. Результатом этой деятельности должен стать план обработки риска. Управление рисками, включая выявление, оценку и обработку связанного с ИИ риска, описано в стандарте ISO/IEC 23894.

Организации должны определить потенциальный ущерб и преимущества, связанные с ИИ-системой, в том числе путём проведения

консультаций с типичными пользователями. В результате данного процесса может быть сформирован набор ценностей, способных в дальнейшем направлять разработку частей и элементов системы, включая функциональные возможности системы, пользовательский интерфейс, документацию и варианты использования. Организациям следует дополнительно изучать и уточнять эти ценности до такой степени, чтобы те могли стать частью требований к системе. Правовые концепции, концепции прав человека, социальной ответственности и защиты окружающей среды могут помочь в уточнении и описании ценностей.

В дополнение к обычно рассматриваемым рискам, связанным с системой, таким как риски для безопасности и защиты персональных данных, план обработки риска также должен предусматривать обработку рисков, связанных с выявленными ценностями.

Прозрачность и подотчётность: Заинтересованные стороны должны обеспечить, чтобы на протяжении всего жизненного цикла документировались такие аспекты, как происхождение данных, достоверность источников данных, усилия по смягчению рисков, реализованные процессы и решения, - с тем, чтобы способствовать всестороннему пониманию того, как получаются результаты ИИ-системы, а также для целей подотчётности.

Затраты и финансирование: Заинтересованные стороны должны прогнозировать затраты на ИИ-систему в течение жизненного цикла и обеспечивать наличие финансирования.

Ресурсы: Заинтересованные стороны должны определить, какие ресурсы требуются для реализации и завершения каждой из стадий жизненного цикла, и обеспечить доступность этих ресурсов при возникновении в них потребности. Следует обратить внимание на данные, которые могут потребоваться для разработки и/или оценки ИИ-

системы. В случае использующей машинное обучение ИИ-системы особое внимание следует уделять обучающим, валидационным и тестовым данным

Реализуемость: Начальная стадия подводит к принятию решения о том, является ли ИИ-система реализуемой. Может быть проведена демонстрация работоспособности концепции с целью определить, соответствует ли система требованиям и целям. В число примеров требований и целей могут входить следующие:

- система решает поставленную проблему;
- система реализует деловую возможность или обеспечивает выполнение миссии;
- система обеспечивает указанные возможности и характеристики.

Если ИИ-система признаётся реализуемой, то заинтересованные стороны могут принять решение о переходе на стадию проектирования и разработки.

6.2.3 Проектирование и разработка

Данная стадия начинается с проектирования и разработки ИИ-системы и завершается, когда ИИ-система готова к прохождению верификации и валидации. На этой стадии, и в особенности перед её завершением, заинтересованные стороны должны обеспечить, чтобы ИИ-система удовлетворяла первоначальным целям, требованиям, а также другим целям, выявленным на начальной стадии. Примерами процессов, которые могут происходить на стадии проектирования и разработки, являются:

Подход: Заинтересованные стороны должны определить общий подход к проектированию, тестированию и подготовке к приёмке и развертыванию ИИ-системы. Выбор подхода может включать рассмотрение того, потребуются ли оборудование и программное

обеспечение; откуда брать компоненты (например, разрабатывать с нуля, покупать имеющееся на рынке оборудование, использовать программное обеспечение с открытым исходным кодом).

Архитектура: Заинтересованные стороны должны определить и задокументировать общую архитектуру ИИ-системы. Процессы выбора архитектуры и подхода взаимосвязаны, поэтому могут потребоваться итерации их поочерёдного выполнения.

Код: Разрабатывается или приобретается программный код для ИИ-системы.

Обучающие данные: ИИ-системы являются воплощением приобретённых знаний. Обработка обучающих данных является фундаментальной частью процесса разработки ИИ-систем, использующих машинное обучение (см. п.5.10).

Обработка риска: Организации должны внедрить процессы и меры контроля и управления, предусмотренные планом обработки риска (см. стандарт ИСО/МЭК 23894).

6.2.4 Верификация и валидация

На стадии верификации и валидации проверяется, что созданная на стадии проектирования и разработки ИИ-система работает в соответствии с требованиями и соответствует поставленным целям.

Примерами процессов, которые могут происходить на стадии верификации и валидации, являются:

Верификация: Как для программного обеспечения, так и для оборудования проводится тестирование с целью проверки функциональных возможностей и выявления ошибок и недочётов. Также может быть проведено тестирование интеграции систем. Можно провести тесты производительности на соответствие времени отклика,

запаздывания и других существенных показателей производительности ИИ-системы установленным требованиям.

Важным аспектом ИИ-систем является необходимость убедиться в том, что возможности ИИ работают так, как предполагается. Это требует комплектования, подготовки и использования тестовых данных. Тестовые данные должны быть отдельными от любых других данных, используемых в ходе проектирования и разработки, а также репрезентативными в отношении тех входных данных, которые, как ожидается, будет обрабатывать ИИ-система.

Приёмка: Заинтересованные стороны признают ИИ-систему функционально завершённой, имеющей приемлемый уровень качества и готовой к развёртыванию.

Мониторинг и пересмотр рисков: Организации должны, в соответствии с планом обработки риска, проводить анализ результатов верификации, тестирования и валидации для того, чтобы знать о приводящим к рискам событиях и условиях (см. ИСО/МЭК 23894).

6.2.5 Развёртывание

На стадии развёртывания ИИ-система устанавливается, выпускается и/или настраивается (конфигурируется) для функционирования в целевом окружении. Примерами процессов, которые могут происходить на стадии развертывания, являются:

Цель: ИИ-системы могут быть разработаны в одном окружении, а затем развёрнуты в другом. Например, система автоматического беспилотного управления транспортным средством может быть разработана в лаборатории, а затем развернута в миллионах автомобилей. Другие типы ИИ-систем могут быть разработаны на устройствах клиентов, а впоследствии развёрнуты в облаке. Для некоторых ИИ-систем важно различать развёртываемые компоненты

программного обеспечения и используемую (этим программным обеспечением) модель, которая может быть развёрнута отдельно. В таких случаях программное обеспечение и модель могут быть развёрнуты независимо друг от друга.

Обработка риска: Организации должны анализировать и совершенствовать процессы и механизмы управления рисками, а также могут скорректировать план обработки риска (см. ИСО/МЭК 23894).

6.2.6 Эксплуатация и мониторинг

На стадии эксплуатации и мониторинга ИИ-система работает и обычно доступна для использования.

Примерами процессов, которые могут происходить на стадии эксплуатации и мониторинга, являются:

Мониторинг: Ведётся мониторинг как нормального функционирования ИИ-системы, так и инцидентов, включая недоступность, сбои во время выполнения и ошибки. Об этих событиях сообщается соответствующим поставщикам ИИ для принятия мер.

Наладка и ремонт: Если ИИ-система не работает, сбоит или работает с ошибками, то может потребоваться проведение ремонтных работ и технического обслуживания системы.

Обновление: Могут проводиться обновления программного обеспечения, модели и аппаратного обеспечения ИИ-системы с целью выполнения новых требований и повышения эффективности, производительности и надёжности.

Поддержка: Пользователям ИИ-системы предоставляется любая необходимая поддержка, которая нужна для успешного использования системы.

Мониторинг и пересмотр рисков: Организации должны вести мониторинг ИИ-системы во время её эксплуатации с целью обеспечить и

повысить качество и эффективность процесса управления рисками (см. ИСО/МЭК 23894).

6.2.7 Непрерывная валидация

Если ИИ-система использует непрерывное обучение, то стадия эксплуатации и мониторинга дополняется стадией непрерывной валидации. На этой стадии всё то время, пока система работает в режиме промышленной эксплуатации, на постоянной основе проводится инкрементальное обучение. Работа ИИ-системы постоянно проверяется на корректность с использованием тестовых данных. В такой ситуации также может потребоваться обновление самих тестовых данных с тем, чтобы сделать их более репрезентативными в отношении текущих эксплуатационных данных и, тем самым, обеспечить более верную оценку возможностей ИИ-системы.

Непрерывное совершенствование управления рисками: Непрерывную валидацию также следует использовать для обеспечения непрерывного совершенствования процессов управления рисками (см. стандарт ИСО/МЭК 23894).

6.2.8 Повторная оценка

После стадии эксплуатации и мониторинга, с учётом результатов работы ИИ-системы, может возникнуть необходимость в прохождении стадии повторной оценки. Примерами процессов, которые могут происходить на стадии повторной оценки, являются:

Оценка результатов работы: Результаты ИИ-системы в ходе её эксплуатации должны быть оценены и сопоставлены с выявленными для неё целями и рисками.

Уточнение целей: Уточнение целей осуществляется, если первоначальные цели не могут быть достигнуты ИИ-системой, или если

по мере накопления опыта эксплуатации системы будет выявлена необходимость в модификации целей.

Уточнение требований: Опыт эксплуатации может показать, что некоторые из первоначальных требований являются в определённых аспектах некорректными, и это может привести к уточнению требований, в рамках которого также возможно появление новых и/или исключение некоторых существующих требований.

Мониторинг и пересмотр рисков: Организации должны вести мониторинг приводящих к рискам событий и условий, в соответствии с тем, как это описано в плане обработки риска (см. стандарт ИСО/МЭК 23894).

6.2.9 Вывод из эксплуатации

В какой-то момент ИИ-система может устареть до такой степени, что её ремонт, исправления и обновления уже не будут способны обеспечить удовлетворение новых требований. Примерами процессов, которые могут происходить на стадии снятия с эксплуатации, являются:

Вывод из эксплуатации и утилизация: Если потребность в ИИ-системе отпала или появился более совершенный подход к построению подобных систем, то ИИ-система может быть выведена из эксплуатации и утилизирована. Этот процесс может охватывать данные, используемые системой.

Замена: Если назначение ИИ-системы продолжает оставаться актуальным, но появился более совершенный подход, то может быть проведена замена ИИ-системы (или её компонентов).

7 Обзор ИИ-систем с функциональной точки зрения

7.1 Общие положения

В настоящем стандарте ИИ-система определяется как техническая система, которая порождает такие конечные результаты, как контент, прогнозы, рекомендации или решения для заданного набора определенных человеком целей. ИИ-системы не способны «понимать»; они нуждаются в осуществляемом человеком выборе проектных решений, проектировании, разработке и надзоре. Степень такого надзора зависит от варианта использования. Как минимум, надзор обычно имеет место во время обучения и валидации. Такой надзор полезен для обеспечения того, что ИИ-система разрабатывается и используется так, как предполагалось, и что её воздействие на заинтересованные стороны надлежащим образом принимается во внимание на протяжении всего жизненного цикла системы.

На рисунке 5 показано функциональное представление ИИ-системы, в которой входные данные обрабатываются с использованием модели для получения выходных результатов. Модель может быть создана либо непосредственно, либо путём обучения на обучающих данных. Пунктирными линиями показаны элементы, специфические для ИИ-систем, использующих машинное обучение.

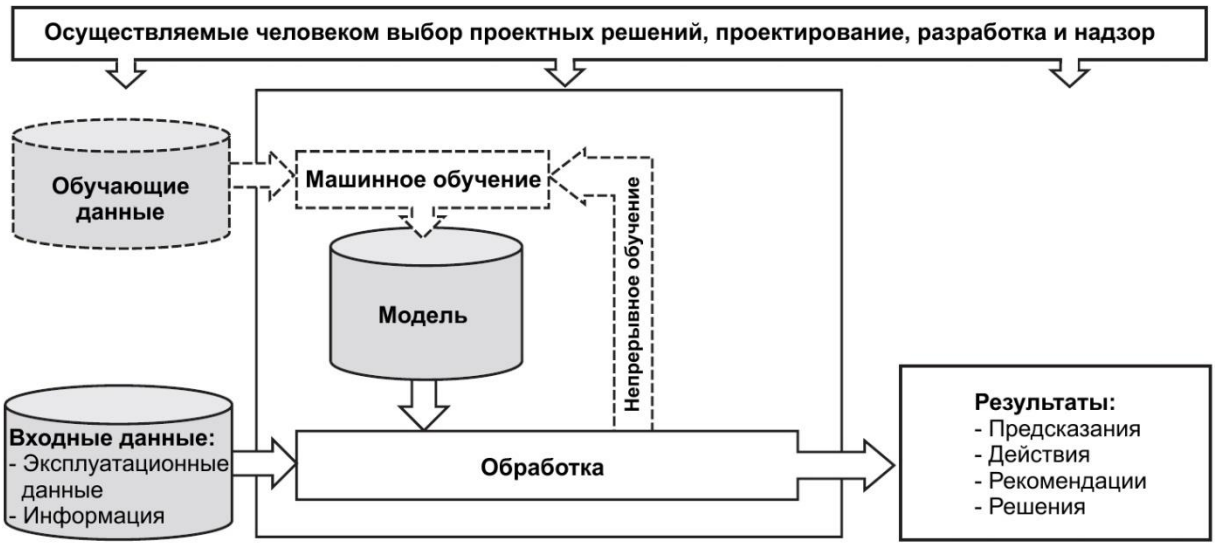

Рисунок 5 — Функциональное представление ИИ-системы

Цель данного представления - дать нетехническое описание того, что ИИ-системы делают для достижения результата. Если говорить коротко, то ИИ-системы содержат модель, которую они используют для порождения прогнозов, а эти прогнозы, в свою очередь, используются для того, чтобы - полностью или частично силами самой системы или с помощью человека - последовательно выдавать рекомендации, решения и выполнять действия.

7.2 Данные и информация

Данные могут поступать на вход находящейся в эксплуатации ИИ-системы, в этаком случае они называются эксплуатационными данными. Может потребоваться проведение предварительной обработки входных данных до того, как они будут направлены в ИИ-систему - например, извлечение соответствующих признаков.

На вход ИИ-системы вместо данных также может поступать информация – обычно в задачах оптимизации, когда единственным

необходимым входом является информация о том, что следует оптимизировать. Некоторые ИИ-системы вообще не требуют никаких входных данных, вместо этого выполняя определённую задачу по запросу (например, создавая синтетическое изображение).

В случае машинного обучения обучающие данные используются для приобретения определённой информации о представляющей интерес области и о задаче, которую следует решить.

При разработке и оценке ИИ-систем данные используются и в иных целях (см. п.5.10).

7.3 Знания и обучение

Модель, используемая ИИ-системой в ходе эксплуатации и для решения задач, представляет собой машиночитаемое представление знаний.

Существует два основных типа таких знаний: декларативные и процедурные:

- декларативные знания – это знания о том, что существует. Такие знания легко облекать в слова (вербализировать) и преобразовывать в утверждения. Например, утверждение «бледная поганка – ядовитый гриб» является декларативным знанием;

- процедурные знания (также известные как ноу-хау) - это знания о том, как что-либо сделать. В контексте ИИ это могут быть модели машинного обучения и другие модели, основанные на подходах, которые включают в себя управление данными и субъективный опыт, полученный от эксперта. Довольно часто их трудно высказать словами (вербализировать). Они транслируются в процедуры. Например, для того, чтобы узнать, является ли гриб ядовитым, можно воспользоваться процедурными знаниями: «Если у вас есть книга о грибах, посмотрите,

сможете ли вы с её помощью определить свой гриб. Если да, то книга даст вам ответ. Если нет, то обратитесь к фармацевту».

Знания имеют различные возможные представления, от неявных до явных.

Знания также могут поступать из различных источников, в зависимости от используемых алгоритмов: они могут уже иметься, их можно приобрести посредством измерений с помощью датчиков и процессов обучения, или же можно использовать комбинацию обоих способов.

Эвристические системы: ИИ-системы, которые не применяют обучение, называются эвристическими. Хорошими их примерами служат классические экспертные системы и системы рассуждений, использующие фиксированную базу знаний. В таких случаях разработчики систем используют человеческие знания для того, чтобы сформулировать разумные правила, определяющие поведение ИИ-системы.

ИИ-системы, использующие машинное обучение: Говорят, что ИИ-системы, включающие процесс обучения, «используют машинное обучение». Обучение включает в себя вычислительный анализ обучающего набора данных с целью выявления закономерностей, создание модели и сравнение результатов полученной модели с ожидаемым поведением. Данный процесс также известен как «тренировка, тренинг» (training). Полученная база знаний представляет собой модель, обученную с использованием математической функции и обучающего набора данных, которая является наилучшей аппроксимацией поведения в заданном окружении.

Непрерывное обучение: ИИ-системы также различаются с точки зрения того, когда и как поступают данные. В некоторых случаях база знаний является статичной и предоставляется с самого начала вместе с

предварительно запрограммированными компонентами системы. В других случаях база знаний изменяется и/или адаптируется с течением времени, при этом информация обновляется в ходе эксплуатации ИИ-системы. Системы машинного обучения можно характеризовать на основании того, когда в их жизненном цикле происходит обучение. Во многих случаях первоначальная фаза обучения позволяет получить некоторое приближение к желаемой целевой функции, и система продолжает использоваться «как есть», без обновления этого внутреннего представления на основе новых примеров. При использовании альтернативного подхода, известного как непрерывное обучение (или обучение на протяжении всего жизненного цикла), обучение распределено во времени; модель обновляется итеративно, по мере того как становятся доступны новые данные. На практике модели, использующие обучение на протяжении всего жизненного цикла, обычно реализуют комбинацию обоих подходов – после первоначальной фазы обучения, на которую приходится основная часть обучения, модель затем уточняется с течением времени на основе новых данных.

7.4 От прогнозов до действий

7.4.1 Общие положения

Результаты обработки ИИ-системой входных данных могут быть различной природы, в зависимости от уровня автоматизации системы. В зависимости от варианта использования ИИ-система может как выдавать только первичные «техническое» результаты (прогнозы), так и предпринимать более эффективные шаги, предлагая или самостоятельно выполняя действия в своём окружении (рекомендации, решения и, наконец, действия).

При классификации ошибочные результаты обычно категорируются как ложноположительные или ложноотрицательные. Ложноположительным результатом является положительный прогноз в случае, когда реальный результат оказывается отрицательным. Ложноотрицательный результат появляется в случае, когда модель ошибочно прогнозирует отрицательный результат. Пользователи ИИ-систем должны понимать последствия ошибочных результатов, включая возможность предвзятых прогнозов. Проблемы такого рода могут являться непосредственным отражением свойств и характеристик инструментов, процессов или данных, используемых для разработки системы.

Ключевым моментом является то, что результаты ИИ-системы могут быть ошибочными. Скорее можно говорить о вероятности результата оказаться правильным, чем об абсолютной правильности. Как разработчики, так и пользователи ИИ-систем должны знать, что подобные системы могут выдавать неправильные результаты, и понимать последствия использования таких неправильных результатов с точки зрения подотчётности.

7.4.2 Прогноз

Термин «прогноз» относится к самому первому результату ИИ-системы на выходе.

ИИ-системы делают прогнозы, применяя модель к новым данным или ситуациям. В примере с принятием решения о выдаче кредита в п.7.4.3, ИИ-система была разработана на основе предыдущих документов о выдаче кредитов. Продолжая пример: когда новый человек подаёт заявку на кредит, его информация передаётся модели, которая затем даёт оценку вероятности того, что данный человек сможет выплатить кредит.

Примечание – При использовании искусственного интеллекта под «прогнозом» не обязательно понимается утверждение о будущем – данный термин относится лишь к результату ИИ-системы на выходе, которым может быть вид цветка на изображении или перевод на другой язык.

7.4.3 Решение

Под «решением» понимается выбор конкретного способа действий, с намерением применить его.

Решения могут быть приняты на основе полученных системой результатов как самой системой, так и людьми. Решения могут быть приняты на основе рекомендаций или же непосредственно на основе прогнозов.

Например, если, согласно прогнозу ИИ-системы, есть существенный риск того, что потенциальный заёмщик не сможет погасить кредит, то оформляющий кредиты сотрудник кредитного учреждения (человек) может проанализировать этот результат вместе с другой информацией о данном заёмщике и о ситуации у кредитора, и затем принять решение об одобрении заявки на кредит. В качестве альтернативы, система сама может дать рекомендацию об одобрении кредита и оценить вероятность того, что это наилучший вариант действий с учётом ожиданий кредитора; и тогда оформляющий кредиты сотрудник может принять решение одобрить кредит, если сочтёт данную вероятность приемлемой. Или же заявка на кредит может быть одобрена автоматически, на основе пороговых значений для принятия системой решения на основе таких рекомендаций.

Человеческое суждение и надзор различными способами вовлекаются в такой процесс принятия решений. Устанавливаемые человеком пороговые значения обычно выбираются с учётом рисков, связанные с автоматизацией принятия решений. Даже когда процесс

принятия решений полностью автоматизирован, люди могут использовать прогнозы для мониторинга получаемых решений.

7.4.4 Действие

За решениями следуют действия, и именно в этом момент результаты ИИ-системы начинают влиять на реальный мир (как физический, так и виртуальный).

Выполнение действия является последним этапом применения информации в ИИ-системе. Например, в примере с принятием решения о выдаче кредита из п.7.4.2, как только кредит будет одобрен, последующие действия могут включать подготовку кредитных документов, получение подписей и выполнение платежей. В случае с роботом, действием может быть выдача приводам робота команд на позиционирование его рук и ладоней. В зависимости от ИИ-системы, действие может происходить в пределах границ ИИ-системы или за их пределами.

8 Экосистема ИИ

8.1 Общие положения

На рисунке 6 экосистема ИИ представлена с точки зрения функциональных уровней. Крупные ИИ-системы полагаются не на одну какую-либо технологию, а, скорее, на сочетание разработанных в разное время технологий. Такие системы могут одновременно использовать различные технологии - например, нейронные сети, символьные модели и вероятностные рассуждения.

Каждый уровень на рисунке 6 использует ресурсы нижележащих уровней для реализации своих функций. Более светлые затенённые

прямоугольники обозначают субкомпоненты уровня или функции. Геометрические размеры уровней и субкомпонентов не отражают их важность.

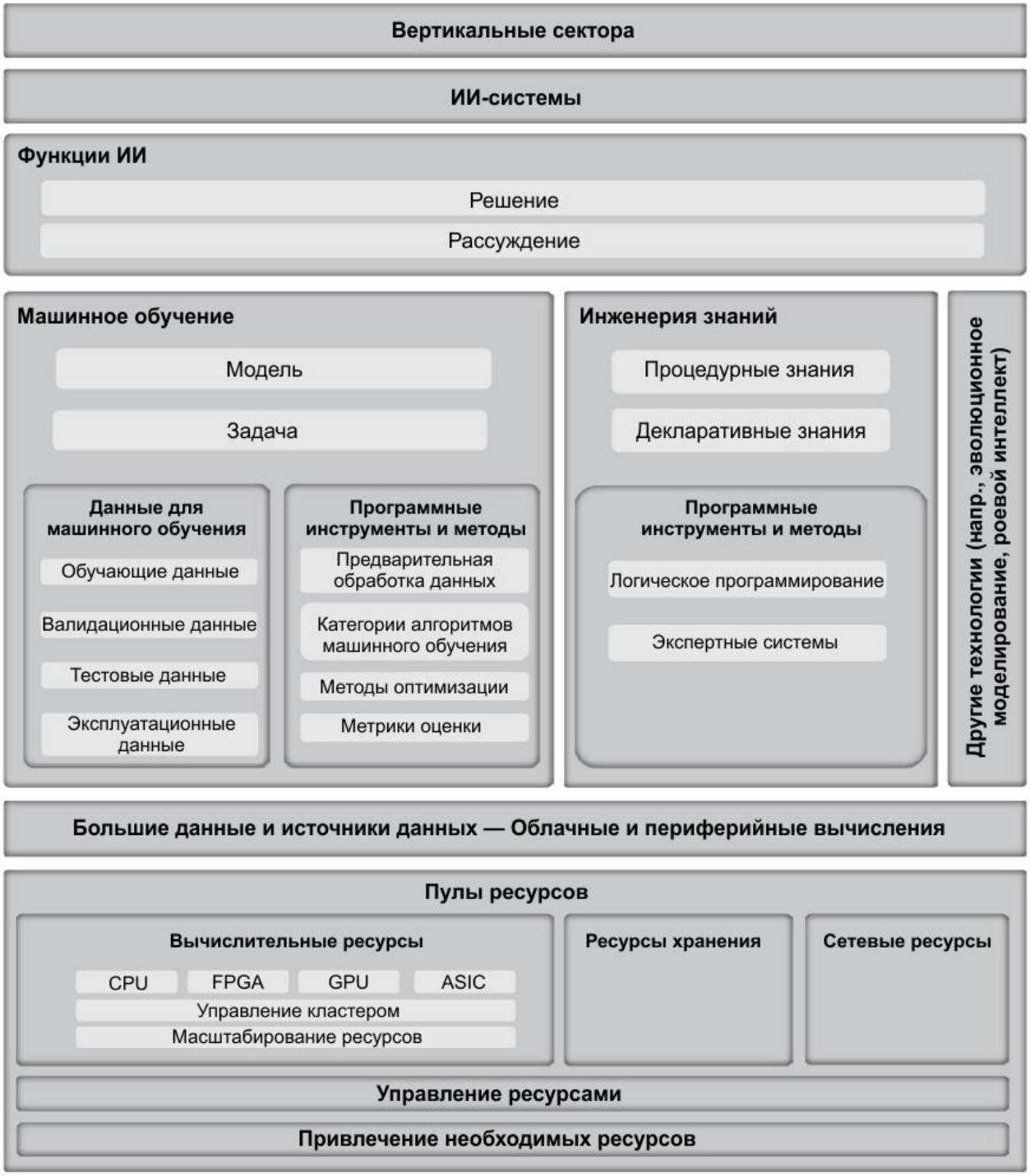

Рисунок 6 — Экосистема ИИ

Создание ИИ-систем продолжает оставаться предметом современных исследований. Между тем использование технологий ИИ становится неотъемлемой частью деятельности во многих сферах, каждая из которых имеет свои потребности, ценности и нормативные правовые ограничения.

Специализированные ИИ-приложения, используемые, например, для компьютерного зрения или обработки естественного языка, сами становятся строительными блоками при создании различных продуктов и услуг. Такие приложения являются движущей силой проектирования специализированных ИИ-систем и, как следствие, устанавливают приоритеты для исследований и разработок.

Технологии ИИ часто требуют использования значительных вычислительных, сетевых ресурсов и ресурсов хранения, например, на этапе обучения ИИ-системы, использующей машинное обучение. Такие ресурсы, как показано на рисунке 6, могут быть эффективно получены с использованием облачных вычислений.

В следующих подразделах описываются основные компоненты показанной на рисунке 6 экосистемы ИИ.

8.2 ИИ-системы

ИИ-системы могут использоваться во многих приложениях и для решения множества различных задач. В разделе 9 описываются примеры приложений с использованием ИИ, таких, как распознавание образов, обработка естественного языка и прогнозная техническая поддержка (predictive maintenance). В разделе 5 перечислены многочисленные типы задач, которые способны решать ИИ-системы.

ИИ-системы следуют универсальному функциональному подходу, когда для создания модели прикладной области информация собирается либо путём прямого встраивания в программный код (с использованием

инженерии знаний), либо путём машинного обучения. Далее закодированная в виде модели информация используется на уровне рассуждений, где вычисляются потенциальные решения; а затем - на уровне принятия решений, где делается выбор между возможными действиями, которые могут привести к цели. Уровень рассуждений включает в себя рассуждения, основанные на представлениях о пространстве, времени и здравом смысле, вычисляемые приложения политики и/или любую иную поддающуюся кодированию форму рассуждений. На уровне принятия решения делается выбор среди возможных действий на основе предпочтений или полезности.

8.3 Функции ИИ

После того, как модель создана, функции ИИ заключаются в выработке прогноза, рекомендации или, в более общем смысле, в принятии решения, которые помогут достичь текущую цель ИИ-системы.

Под «рассуждением» понимается только лишь применение имеющихся в текущей ситуации данных к модели, и задание модели вопроса о том, какие имеются возможные варианты.

Примерами технологий, реализующих различные формы рассуждений, служат планирование, байесовский вывод, автоматические системы доказывания теорем, пространственные и временные рассуждения и рассуждения на основе онтологий.

Система к тому же должна решить, какой из этих возможных вариантов, которые, вероятно, позволят достичь цели, является лучшим.

Здесь в игру вступают предпочтения и полезность: автоматизированное такси будет максимизировать благополучие клиента, а программа игры в покер будет максимизировать свою прибыль.

8.4 Машинное обучение

8.4.1 Общие положения

Машинное обучение – это процесс, использующий вычислительные методы для того, чтобы дать системам возможность обучаться на данных или опыте. В нём применяется ряд статистических методов для поиска закономерностей в имеющихся данных, а затем эти закономерности используются для создания прогнозов на основе эксплуатационных данных.

В традиционном программировании разработчик программного обеспечения определяет логику решения поставленной задачи, задавая точно определённые шаги вычислений с использованием языка программирования. По контрасту, логика модели машинного обучения частично зависит от данных, используемых для обучения модели. Таким образом, в случае машинного обучения необходимые для решения задачи вычисления или шаги не определяются априори.

Кроме того, в отличие от традиционного программирования, модели машинного обучения могут совершенствоваться с течением времени, не требуя при этом переписывания – это делается посредством повторного обучения с использованием дополнительных новых данных и с помощью методов оптимизации параметров модели и выделяемых в данных признаков.

8.5 Инженерия знаний

8.5.1 Общие положения

При использовании экспертами-людьми подхода, основанного на инженерии знаний, характер обработки зависит исключительно от экспертных знаний разработчика и понимания им задачи. Знания

приобретаются не через обучение на основе данных, а путем жёсткого кодирования разработчиком в ИИ-системе знаний экспертов в предметной области.

Существует два основных типа знаний: декларативные и процедурные. Более подробную информацию об обоих типах знаний см. в п.7.3.

8.5.2 Экспертные системы

Как подразумевает сам термин, экспертная система - это ИИ-система, которая накапливает, комбинирует и объединяет предоставленные экспертами-людьми знания в предметной области с целью логического вывода решений поставленных задач.

Экспертная система состоит из базы знаний, механизма логического вывода и пользовательского интерфейса. В базе знаний хранятся декларативные знания о предметной области, охватывающие как фактическую, так и эвристическую информацию. Механизм логического вывода содержит процедурные знания: набор правил и методологию рассуждений. Он комбинирует предоставленные пользователем факты с информацией из базы знаний.

Логический вывод делается с использованием предопределенных правил, согласованных с экспертом, и с оценками логических утверждений. В число задач, которые могут быть решены с использованием экспертных систем, входят задачи классификации, диагностики, мониторинга и прогноза.

8.5.3 Логическое программирование

Логическое программирование – это форма программирования, основанная на языках программирования, позволяющих представлять логические утверждения, записанные на языке формальной

(математической) логики. Примером языка логического программирования является Prolog.

С точки зрения ИИ формальная логика была важной областью исследований. Многие виды формальной логики нацелены на моделирование человеческих рассуждений в различных ситуациях. Логическое программирование обеспечивает среду для реализации таких моделей человеческих рассуждений. Агенты ИИ должны быть способны воспроизводить различные виды рассуждений четко определённым, прозрачным и объяснимым образом.

Логическое программирование с декларативными утверждениями в сочетании с эффективной обработкой естественного языка может создать для агента ИИ возможности для того, чтобы рассуждать по аналогии, делать выводы и обобщения об объектах и окружении.

Пример: Apache Jena [19] - это среда семантической «паутины», которая поддерживает механизм логического вывода.

8.6 Большие данные и источники данных - облачные и периферийные вычисления

8.6.1 Большие данные и источники данных

Все системы машинного обучения используют данные. Эти данные могут принимать различные формы. В некоторых случаях используемые системами машинного обучения данные являются «большими данными». Соответствующий уровень на рисунке 6 представляет источники, форматы и типичные методы оперирования больших данных, вне зависимости от способов их использования. Данный подраздел детально описывает основные компоненты, показанные на рисунке 7.

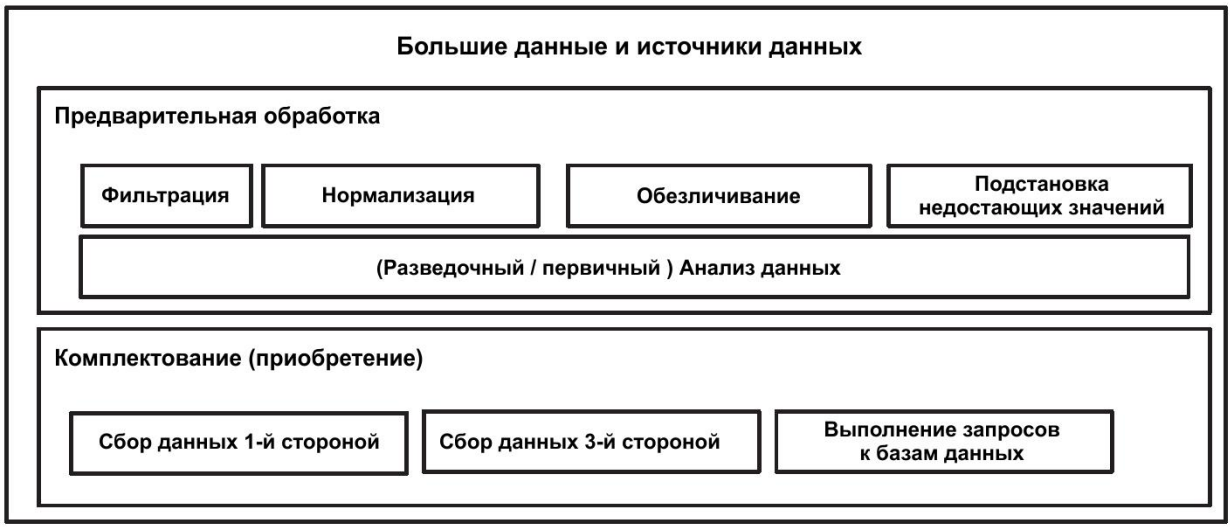

Рисунок 7 — Большие данные и источники данных

Большие данные – это обширные наборы данных, характеристики которых с точки зрения объёма, разнообразия, скорости и вариативности требуют применения специализированных технологий и методов для их обработки и получения от них отдачи. Так, например, были разработаны технологии специально для обеспечения возможности распределённой обработки больших наборов данных с использованием вычислительных кластеров и простых моделей программирования. Кроме того, технологии хранения и баз данных были разработаны специально для управления большими объемами данных, которые могут формироваться из других массивов данных большого объёма.

Большие данные стали важны ввиду того, что организации увеличили широту и глубину сбора данных, из-за чего потребовались специализированные технологии и методы для извлечения знаний.

Дополнительную информации о больших данных см. в ИСО/МЭК 20546 и ИСО/МЭК 20547-3.

Функционирование многих ИИ-систем невозможно без больших данных, которые широко применяются в ИИ. Доступность больших наборов неструктурированных данных в различных областях применения

приводит к получению новых знаний в результате применения таких методов ИИ, как интеллектуальный анализ данных или распознавание образов. Доступность огромных объемов данных для обучения приводит к появлению более совершенных моделей машинного обучения, которые могут быть использованы в широком спектре приложений.

Данные может приобретать та же организация, которая их использует (сбор данных первой стороной). Например, предприятия розничной торговли используют данные о транзакциях, которые они получают из принадлежащих им систем кассовых терминалов. Данные также могут быть приобретены третьими сторонами, такими как научно-исследовательские организации и иные поставщики данных, которые собирают данные, а затем продают их или обмениваются ими с другими организациями, непосредственно использующими эти данные. Кроме того, данные могут быть получены посредством выполнения запросов и слияния данных из различных наборов и баз данных как первой, так и третьей стороны.

Данные могут поступать из многих источников, таких как:

- оплата покупок и другие транзакции;
- опросы и обследования;
- статистические исследования;
- задокументированные наблюдения;
- датчики (сенсоры);
- изображения;
- аудиозаписи;
- документы;
- взаимодействие с системами.

8.6.2 Облачные и периферийные вычисления

Облачные вычисления – это парадигма для обеспечения сетевого доступа к масштабируемому и гибкому пулу совместно используемых физических и/или виртуальных ресурсов с системой самообслуживания и администрированием по требованию, см. стандарты ИСО/МЭК 17788 и ИСО/МЭК 17789.

Облачные вычисления обычно ассоциируются с крупными централизованными центрами обработки данных, способными обеспечить очень большие вычислительные ресурсы для обработки и хранения данных. Такие большие возможности могут играть решающую роль на некоторых стадиях жизненного цикла ИИ-систем, особенно при обработке больших наборов данных для обучения ИИ-систем и создания используемых ими моделей.

Периферийные вычисления - это распределенные вычисления, в которых обработка и хранение данных осуществляются на или около периферии сети, при этом степень близости к периферии определяется требованиями системы. Периферия сети - это граница между соответствующими цифровыми и физическими сущностями, проходящая через подключённые к сети датчики и исполнительные устройства (см. стандарт ИСО/МЭК 23188).

Концепция периферийных вычислений в значительной степени касается размещения и функционирования программных компонентов и хранения данных. В тех случаях, когда программные компоненты (например, связанные с ИИ-системами) имеют дело с устройствами интернета вещей (датчиками и приводами), часто существует потребность в минимизации задержек и в выдаче результатов в рамках существенных ограничений по времени (как часто говорят, в режиме реального времени); и/или потребность в обеспечении жизнеспособности, чтобы система могла по-прежнему функционировать

в случае перебоев со связью; и/или потребность в защите персональных данных физических лиц, полученных от периферийных устройств. Для достижения этих целей может потребоваться, чтобы обработка и хранение данных выполнялись на периферии сети или поблизости от неё. Дополнительную информацию об этом см. в ИСО/МЭК 23188.

Важно, однако, понимать, что облачные вычисления могут быть развёрнуты во многих местах распределённой вычислительной среды, в том числе в таких, которые не являются централизованными и которые находятся вблизи периферии сети. В такой форме облачные вычисления могут предложить гибкое и динамичное развёртывание как для программного обеспечения, так и для данных, используя виртуализированную обработку и виртуализированное хранение данных в сочетании с объединением ресурсов в пул и с быстрой эластичностью и масштабируемостью, - создавая тем самым условия для адекватного размещения и функционирования компонентов ИИ-систем.

Обычно системы периферийных вычислений комбинируются с централизованными системами для создания законченных решений, что позволяет воспользоваться возможностями систем обоих типов.

Три основных проектных решения для систем на основе машинного обучения соединяют в себе облачные и периферийные вычисления: это обучение модели в облаке, обучение модели на периферии, обучение модели в облаке и на периферии.

а) Облачные сервисы могут быть использованы в качестве централизованной платформы для обучения моделей машинного обучения (см. рисунок 8). Ввиду ограниченных ресурсов периферийных устройств, требовательные к вычислительным ресурсам и ресурсам хранения задачи, связанные с обучением, валидацией и поддержкой моделей, выполняются с использованием облачной инфраструктуры. Обученная модель развертывается, применяется и, при необходимости,

обновляется на периферийных устройствах. Данные с периферийных устройств могут быть затем использованы при обучении или, как в случае обучения с подкреплением, для организации обратной связи по качеству модели.

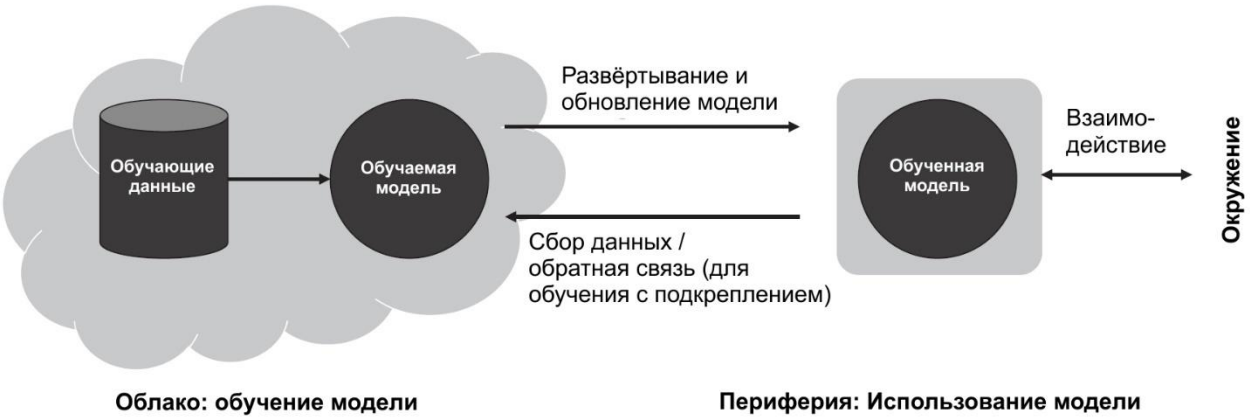

Рисунок 8 — Пример обучения модели в облаке

Примерами приложений, использующих такое проектное решение, являются обнаружение атак на пограничные маршрутизаторы (интеллектуальные брандмауэры), обнаружение и профилактика неисправностей в приложениях для управления производственными процессами (профилактическое техническое обслуживание) и распознавание дорожных знаков самоуправляемыми автомобилями.

б) В случае, когда централизованный подход не является оптимальным для обучения персонализированных моделей или моделей, используемых в специфических условиях применения, может применяться другая схема (см. рисунок 9). В её основе лежит обучение модели непосредственно на периферийных устройствах (при условии, что на них имеются достаточные ресурсы).

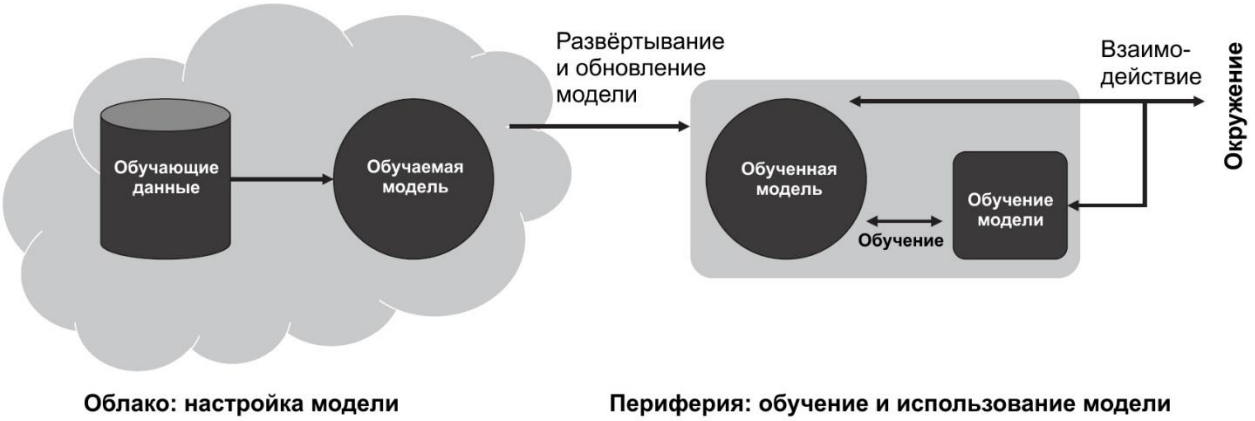

Рисунок 9 — Пример обучения модели на периферии

При таком проектном решении только исходная (типовая) модель настраивается и обучается в облачной среде. Контекстуализированное или персонализированное обучение выполняется на периферии с использованием реальных данных. Этот вариант обучения модели является наиболее подходящим для полностью автоматических систем, в которых применяются такие методы машинного обучения, как обучение без учителя или обучение с подкреплением.

в) При гибридном подходе обучение модели проводится как в облаке, так и на периферии (см. рисунок 10). Это может быть необходимо в случаях, когда проектное решение системы включает периферийные устройства. В некоторых случаях облачные сервисы используются для подготовки исходной обученной модели, которая затем развертывается на периферии. В других случаях периферийные системы обучают их локальные модели на основе своих локальных данных, не передавая данные друг другу и облачным сервисам. Облачные сервисы также могут выступать в качестве сервера параметров, осуществляя синхронизацию обновлений моделей различных периферийных систем и возвращая затем синхронизированные обновления модели в периферийные системы для обновления их индивидуальных моделей.

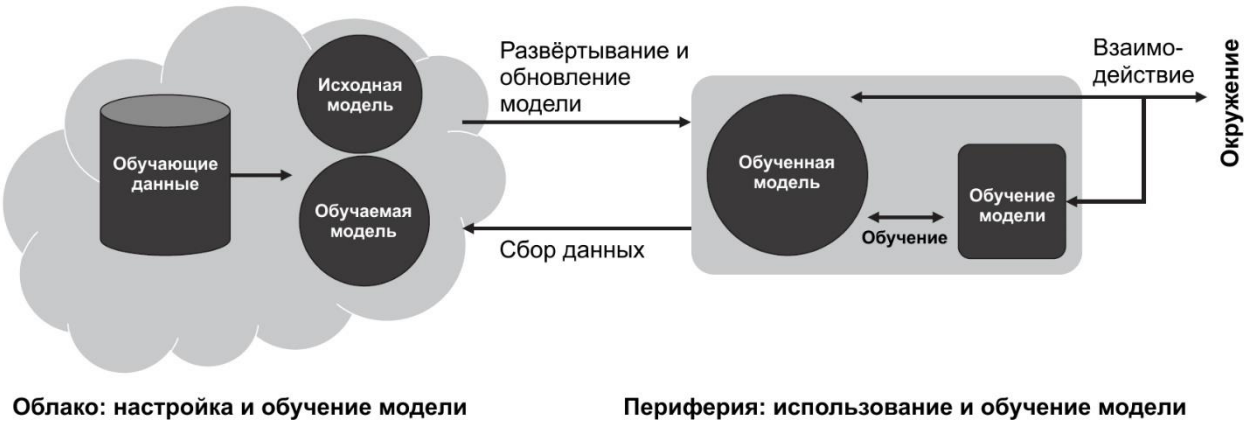

Рисунок 10 — Пример обучения модели в облаке и на периферии

Примером применения такого проектного решения является сервис сбора пространственных данных (например, сервис сбора изображений различных участков местности с беспилотных летательных аппаратов - дронов, или сервис сбора данных с бытовых устройств). Данный подход даёт возможность обеспечить более высокое качество обслуживания благодаря использованию обновляемых обучаемых моделей вместо исходных обученных моделей.

Можно рассмотреть ещё один гибридный подход, включающий загрузку модели. В этом случае обученная на периферии модель посылается в облачный репозиторий, откуда – если у неё показатели производительности лучше, чем у предыдущей обученной модели – она рассылается другим периферийным системам, функционирующим в такой же или аналогичной среде. Данный подход может применяться в трансферном обучении и в методах сжатия моделей. Примером трансферного обучения может служить ситуация, когда обученная модель для распознавания номеров домов при просмотре изображений улиц может использоваться для распознавания рукописных чисел. Исходная модель либо модель, уже обученная для решения конкретной проблемы, может быть применима для решения аналогичных задач. В случае периферийных устройств с меньшей вычислительной мощностью

также могут использоваться методы сжатия модели. Модель, полностью обученная в среде располагающего обильными вычислительными ресурсами облачного сервиса, может быть сжата перед её использованием в периферийной системе с меньшими ресурсами.

8.7 Пулы ресурсов

8.7.1 Общие положения

На рисунке 6 показаны ресурсы, необходимые для поддержки экосистемы ИИ. Для поддержки ИИ-систем крайне важны как вычислительные и сетевые ресурсы, так и ресурсы для хранения данных.

Разработка и развёртывание ИИ-систем может происходить на ресурсах разного масштаба, начиная от централизованных облачных сервисов и локальных центров обработки данных и заканчивая серверами (или кластерами серверов), периферийными вычислительными системами, мобильными устройствами и устройствами интернета вещей. Некоторые из этих систем могут располагать ограниченными ресурсами в плане вычислительной мощности, объёмов хранения данных, а также пропускной способности сети и сетевой задержки. Это в особенности относится к системам и устройствам на периферии. Вычислительные ресурсы ИИ-систем могут быть представлены в виде любой конфигурации отдельных или многочисленных графических процессоров, нейронных процессоров, центральных процессоров и процессоров других типов, входящих в состав как одной системы, так и нескольких систем, которые могут быть объединены в вычислительные кластеры.

Потребности ИИ-систем в вычислительных ресурсах могут варьироваться в зависимости от того, используется ли машинное или глубокое обучение, а также в зависимости от типов рабочей нагрузки

(например, обучение и логический вывод с использованием различных топологий). Вследствие этого могут потребоваться гетерогенные вычислительные решения, соответствующие конкретной рабочей нагрузке и ИИ-системе. Например, аппаратные ускорители (GPU, NPU, FPGA, DSP, ASIC и др.) могут быть использованы для вычислительно интенсивных рабочих нагрузок ИИ-систем, таких как обучение определенных топологий нейронных сетей.

Чтобы удовлетворить потребности разнообразных ИИ-систем, нужно в ходе привлечения необходимых ресурсов поддерживать возможность автоматического управления ресурсами, включая выделение ресурсов по требованию и координацию использования гетерогенных ресурсов (например, выделения локальных, облачных и периферийных ресурсов).

8.7.2 Специализированные интегральные схемы

Специализированные интегральные схемы ASIC – это вид интегральных схем, специализированных под конкретное применение. Их использование является одним из вариантов обеспечения специфических для искусственного интеллекта функциональных возможностей.

Схема ASIC может быть изготовлена и настроена как ускоритель, предназначенный для ускорения процесса ИИ посредством предоставления таких функциональных элементов и возможностей, как специализированные, параллельные работающие блоки умножения с накоплением, оптимизированное распределение памяти и арифметика пониженной точности. Схема ASIC также может быть сконфигурирована как сопроцессор, выполняющий для задач ИИ функции предварительной или последующей обработки данных - например, для кадрирования и

изменения размера изображений, их преобразования, подавления шума или слияния данных от распознанных изображений.

В отличие от универсальных процессоров общего назначения (таких, как центральные и графические процессоры), схемы ASIC обычно проектируются, производятся и используются только для конкретных способов применения, таких как реализация конкретных структур нейронной сети. Схемы ASIC обеспечивают более высокие вычислительные возможности для ИИ при меньших пространственных размерах, более низкой стоимости и сниженном потреблении энергии.

Схемы ASIC дают возможность реализовать ИИ в устройствах с ограниченными габаритами и возможностями источников питания, таких как мобильные телефоны. Схемы ASIC также позволяют использовать ИИ в устройствах интернета вещей, применяемых в различных областях, таких как промышленное производство, здравоохранение, безопасность или технологии «умного дома».

9 Предметные области ИИ

9.1 Компьютерное зрение и распознавание образов

В настоящем стандарте компьютерное зрение определяется как «способность функционального компонента получать, обрабатывать и интерпретировать данные, представляющие изображения или видеосигналы» (п.3.7.1). Компьютерное зрение тесно связано с распознаванием образов, т.е. с обработкой цифровых изображений. Визуальные данные обычно проступают от цифрового датчика изображения как результат оцифровки аналогового изображения путём сканирования или же от иного устройства ввода изображений. Для целей

данного стандарта под цифровыми изображениями понимаются как статические, так и подвижные варианты изображений.

Цифровые изображения существуют как матрицы чисел, представляющие цвета или градации серого цвета в захваченном изображении, а в других случаях – как наборы векторов. Цифровые изображения могут включать метаданные, которые описывают связанные с ними характеристики и атрибуты. Цифровые изображения могут быть сжаты для экономии места хранения и повышения производительности при передаче в цифровых сетях.

Ниже приведены примеры приложений ИИ на основе компьютерного зрения и распознавания образов:

- выявление конкретных образов в наборе изображений (например, изображений собак в наборе изображений животных);
- самоуправляемые автомобили: обнаружение и идентификация автоматизированными транспортными средствами дорожных знаков, сигналов светофоров и объектов;
- медицинская диагностика: выявление заболевания и аномалий при анализе медицинских изображений;
- контроль качества (например, выявление дефектных деталей на сборочной линии);
- распознавание лиц.

В число фундаментальных для компьютерного зрения задач входят получение изображения, повторная дискретизация, масштабирование, снижение уровня шума, повышение контраста, извлечение признаков, сегментация, обнаружение объектов и классификация.

Существует несколько подходов, которые могут быть использованы для выполнения задач компьютерного зрения в ИИ-системах. В последние годы стали популярными глубокие свёрточные нейронные сети (см. п.5.12.1.4) ввиду их высокой точности в задачах классификации

изображений и их показателей производительности в задачах обучения и прогнозирования.

9.2 Обработка естественного языка

9.2.1 Общие положения

Обработка естественного языка - это обработка информации, основанная на понимании естественного языка и/или генерации естественного языка. Данный термин охватывает анализ естественного языка и его генерацию, в форме текста или речи. Используя возможности обработки естественного языка, компьютеры могут анализировать написанный на человеческом языке текст и выделять в нём понятия, сущности, ключевые слова, отношения, эмоции, настроения и другие характеристики, тем самым давая пользователям возможность извлекать из контента знания и представления. Располагая этими возможностями, компьютеры также могут генерировать текст или речь для общения с пользователями. Любая система, которая способна воспринимать и обрабатывать естественный язык (в текстовой или речевой форме) в качестве входных или выходных данных, использует компоненты обработки естественного языка. Примером подобной системы является автоматизированная система бронирования билетов авиакомпании, которая может принимать звонки от клиентов и бронировать для них рейсы. Такая система нуждается в компоненте понимания естественного языка и компоненте генерации естественного языка.

Ниже приведены другие примеры приложений ИИ, основанных на обработке естественного языка:

- распознавание рукописного текста (например, преобразование рукописных заметок в цифровую форму);

- распознавание речи (например, понимание смысла того, что сказал человек);

- выявление спама (например, использование значения слов в сообщении электронной почты для того, чтобы установить, можно ли это сообщение отнести к нежелательным);

- цифровые персональные помощники и онлайновые виртуальные собеседники (чат-боты), которые могут использовать понимание и генерацию естественного языка (включая распознавание и генерацию речи) для организации речевых пользовательских интерфейсов;

- реферирование;

- генерация текста;

- поиск по контенту.

Обработка естественного языка также используется во многих прикладных системах, таких как чат-боты, системы контекстной рекламы, системы перевода речи и системы электронного (дистанционного) обучения.

9.2.2 Компоненты обработки естественного языка

9.2.2.1. Общие положения

Компоненты обработки естественного языка (NLP-компоненты) решают разные задачи. Наиболее распространёнными из них являются следующие:

Понимание естественного языка (NLU): NLU-компонент преобразует текст или речь во внутреннее описание, которое должно передавать семантику исходного материала. Трудности возникают из-за внутренне присущей естественным языкам неоднозначности: слова и предложения по своей природе неоднозначны по смыслу и, следовательно, результат NLU подвержен ошибкам.

Генерация естественного языка (NLG): NLG-компонент преобразует внутреннее описание в текст или речь, которые понятны человеку. Выполнение этой задачи может включать подбор слов и формулировок с тем, чтобы результат казался пользователю более естественным.

Морфологическая разметка (POS): POS-компонент морфологической разметки используется для категоризации каждого слова на входе как грамматического объекта: является ли это слово существительным, прилагательным, глаголом и т.д. На POS-разметку также оказывают влияние многозначность и многовариантность (полисемия) естественного языка.

Распознавание именованных сущностей (NER): NER-компонент стремится распознать денотационные (понимаемые буквально) наименования лиц, мест, организаций или иных сущностей и соответствующим образом разметить последовательности слов в потоке текста или речи. В зависимости от сущности, может быть извлечено большее количество информации. Например, для людей может быть полезно установить их должность или функцию.

Ответы на вопросы: Компонент ответов на вопросы стремится дать наиболее подходящий ответ на заданный человеком вопрос. Пользователь спрашивает что-либо на естественном языке, и система даёт ему ответ также на естественном языке.

Машинный перевод: Компонент машинного перевода автоматически переводит контент на естественном языке с одного языка на другой. Это может быть преобразование текста в текст, речи в текст, речи в речь или текста в речь. Трудности возникают как из-за неоднозначности, когда слово имеет несколько значений, так и по другим причинам, таким как наличие отсылок между предложениями или внутри

них или не высказанные явным образом намерения. Во многих случаях возможны несколько вариантов перевода.

Оптическое распознавание символов (OCR): OCR-компонент стремится преобразовать представленные в виде изображений текстовые документы (возможно, отсканированные в графические образы) в цифровое кодированное представление их контента: текста, таблиц, цифр, заголовков и их взаимосвязей.

Извлечение взаимосвязей: Компонент извлечения взаимосвязей решает задачу выявления и извлечения связей между именованными сущностями и даже между любыми сущностями в потоке входных данных. Например, такой компонент может выявить в поданном на вход тексте о фильмах то, что «Аль Пачино» «снимался в ведущей роли» в фильме «Серпико».

Извлечение информации (IR): Компонент извлечения информации стремится удовлетворить информационные потребности пользователя посредством выполнения поиска по массиву неструктурированного контента. Отражающий потребность пользователя в информации поисковый запрос алгоритмически сопоставляется с каждым элементом в массиве, чтобы предсказать релевантность этого элемента с точки зрения пользовательской информационной потребности. Результат работы данного компонента обычно выдаётся пользователю в виде списка отобранных элементов, ранжированных в порядке уменьшения их релевантности. Компоненты извлечения информации могут быть разработаны для различных естественных языков и для широкого спектра типов представления информации, включая текст в свободном формате, полуструктурированные документы, структурированные документы, аудиозаписи, изображения и видеозаписи.

Анализ тональности (настроений): Компонент анализа тональности стремится к выявлению и категоризации с помощью

вычислительных методов мнений, выраженных во фрагменте текста, речи или изображения. Этот процесс также известен как интеллектуальный анализ мнений. Примерами субъективных аспектов могут служить позитивные или негативные чувства.

Автоматическое реферирование: Компонент автоматического реферирования стремится в более краткой форме передавать содержащуюся в элементе контента важную информацию, используя для этого один из двух подходов (или их комбинацию). Первый подход – это квазиреферирование (extractive summarization), когда из исходного контента отбирается ключевой релевантный контент, чтобы создать сокращённую версию. Второй подход – это абстрактное реферирование (abstractive summarization), стремящееся синтезировать новый, более короткий текст, который передаёт релевантную информацию. Абстрактное реферирование взаимосвязано с генерацией естественного языка.

Управление диалогом: Компонент управления диалогом помогает управлять серией взаимодействий между пользователем и системой, стремясь сделать работу пользователя более удобной за счёт организации этих взаимодействий в форме, напоминающей разговор на естественном языке. В управлении диалогами используется ряд подходов, в том числе декларативные правила, определяющие реакцию на конкретные входные триггеры, и подходы на основе машинного обучения. Управление диалогом может использовать взаимодействие в текстовой форме, например, для обеспечения более удобного общения с компонентами ответов на вопросы. Компонент управления диалогом также может быть интегрирован с компонентами распознавания и синтеза речи для поддержки приложений в персональных помощниках, агентах онлайнового обслуживания клиентов или при использовании роботов для персонального ухода.

9.2.2.2. Машинный перевод

Машинный перевод – это задача обработки естественного языка, при выполнении которой компьютерная система используется для автоматического перевода текста или речи с одного естественного языка на другой.

В общем случае процесс перевода при выполнении его человеком осуществляется за два шага. Первый шаг заключается в расшифровке смысла материала на исходном языке. На втором шаге этот смысл повторно кодируется на целевом языке. Этот процесс требует глубоких знаний в области грамматики, семантики, синтаксиса, идиом, культурного фона и в других областях.

В числе технических проблем, с которыми сталкивается машинный перевод, можно назвать многозначность смысла слов, зависимость от контекста, грамматические различия и языки, использующие иероглифическое письмо. Было разработано много подходов к машинному переводу, в том числе подходы, основанные на правилах, примерах, статистических закономерностях, нейронных сетях или их комбинациях.

В последние годы для выполнения машинного перевода использовались нейронные сети, что привело к поразительным улучшениям в плавности и точности перевода. С целью достижения высокой степени точности соответствующая модель посредством глубокого обучения может быть обучена и настроена под выражения, специфические для области применения.

9.2.2.3. Синтез речи

Система, которая преобразует текст на естественном языке в речь, называется системой синтеза (генерации) речи.

В общем случае процесс синтеза речи включает три этапа: 1) анализ, 2) моделирование, и 3) синтез. Естественность и разборчивость

являются важными характеристиками системы синтеза речи. Естественность показывает, насколько близок результат к человеческой речи, в то время как разборчивость говорит о том, насколько легко людям понять синтезированную речь. Системы синтеза речи обычно стараются максимизировать обе эти характеристики.

Для синтеза речи применяются различные подходы, включая конкатенативный синтез (concatenation synthesis), формантный синтез (formant synthesis), артикуляторный синтез (articulatory synthesis), синтез на основе скрытых марковских моделей (HMM-based synthesis), аддитивный синтез на основе синусоподобных волн (sinewave synthesis) и синтез с использованием глубоких нейронных сетей (DNN). Каждый подход имеет свои сильные и слабые стороны. Некоторые синтезаторы речи на основе глубоких нейронных сетей позволяют получать результаты, приближающиеся по своему качеству к голосу человека.

9.2.2.4. Распознавание речи

В данном стандарте распознавание речи определяется как преобразование функциональным компонентом речевого сигнала в представление содержания речи. Оцифрованная речь - это вид последовательных данных, поэтому методы, способные обрабатывать данные, ассоциированные с интервалом времени, также могут быть использованы и для обработки фонем речи.

Для распознавания речи применяется несколько подходов на основе нейронных сетей. Один из них предусматривает использование нейронной сети с архитектурой долгой краткосрочной памяти (LSTM-сети) [18]. Этот метод позволяет обучать нейронную сеть и развертывать её в качестве решения для распознавания речи, не требуя комбинирования с другими процессами, такими как скрытые марковские модели; и обеспечивает приемлемые показатели производительности при распознавании.

Ниже приведены примеры приложений ИИ на основе распознавания речи:

- речевые командные системы;
- «цифровая» диктовка;
- персональные помощники.

9.2.2.5. Ответы на вопросы

Системы ответа на вопросы дают возможность вводить в них большое количество страниц текста и применяют технологию ответа на вопросы, чтобы дать ответ на вопросы, сформулированные людьми на естественном языке. Данный подход позволяет людям «спрашивать» и получать почти мгновенные ответы на сложные вопросы. В комбинации с другими интерфейсами прикладного программирования и передовыми методами аналитики, технология ответа на вопросы отличается от традиционного поиска по ключевым словам тем, что обеспечивает пользователю более интерактивное взаимодействие.

9.3 Интеллектуальный анализ данных

Под «интеллектуальным анализом данных» понимается применение алгоритмов для выявления в данных достоверной, новой и полезной информации. Интеллектуальный анализ данных приобрёл известность в конце 1990-х годов, и было признано, что он отличается от известных ранее статистических методов. Традиционная статистика основное внимание обращала на сбор данных, являющихся необходимыми и достаточными для окончательного ответа на конкретный вопрос. Интеллектуальный анализ данных обычно применялся в рамках повторного использования данных с целью нахождения приблизительных ответов или имеющих место с определенной вероятностью совпадений с заданными образцами.

Интеллектуальный анализ данных рассматривается как этап алгоритмического моделирования в полном процессе извлечения знаний из данных. Опираясь на опыт ранних усилий в области интеллектуального анализа данных, отраслевой консорциум смог подробно описать все шаги интеллектуального анализа данных в отраслевом стандарте CRISP-DM, опубликованном в 2000 году [28]. Интеллектуальный анализ данных охватывает ряд методов и подходов, включая деревья решений, кластеризацию и классификацию. С появлением в середине 2000-х годов технологий работы с большими данными стало уже невозможно отделять применение алгоритмов от хранения данных, а тщательное формирование выборок уступило место скоростной обработке больших массивов данных. Эти изменения привели к тому, что процесс жизненного цикла извлечения знаний из данных по новой версии «больших данных» стал рассматриваться как деятельность в рамках науки о данных. Несмотря на то, что «извлечение знаний из данных» и «обнаружение знаний» являются распространёнными терминами в сфере ИИ, на деле тот результат, который выдаёт компьютер, представляет собой информацию, а не знания.

9.4 Планирование

Планирование является одной из дисциплин искусственного интеллекта. Оно является критически важным для отраслевых приложений и важным для многих сфер деловой деятельности, таких как управление рисками, здравоохранение, промышленные роботы для совместной работы с человеком (коллаборативные роботы, коботы), кибербезопасность, когнитивные помощники и оборона.

Планирование позволяет машине автоматически находить процедурную последовательность действий, направленных на

достижение определённых целей, при одновременной оптимизации определённых показателей производительности. С точки зрения планирования, система находится в определённом состоянии. Выполнение действия может изменить состояние системы, а последовательность действий, предложенная при планировании, может перевести систему из исходного состояния ближе к целевому состоянию.

10 Применение ИИ-систем

10.1 Общие положения

Поскольку ИИ-системы способны оказывать помощь в процессах принятия решений, а в ряде случаев их полностью автоматизировать, давать рекомендации и помогать в автоматизации определённых задач - они находят применение в различных отраслях, включая следующие:

- сельское хозяйство и фермерская деятельность;
- автомобилестроение;
- банковские и финансовые технологии;
- оборона;
- образование;
- энергетика;
- здравоохранение;
- законодательство и право;
- производство;
- СМИ и развлечения;
- смешанная реальность (включает дополненную реальность и интерактивные возможности для взаимодействия);
- государственный сектор;

- розничная торговля и маркетинг;
- безопасность;
- космические технологии;
- телекоммуникации.

Примеры использования ИИ представлены в подразделах с 10.2 по 10.4.

10.2 Выявление мошенничества

Под мошенничеством понимается использование обмана с целью извлечения прибыли. Мошенничество проявляется во многих областях и в различных формах, включая следующие:

- поддельные деньги и документы;
- украденные кредитные карты и документы;
- частная переписка, такая как электронная почта;
- поддельные или украденные идентификационные данных.

Ниже приведены примеры применения ИИ для выявления случаев мошенничества:

- выявление мошеннических случаев списания денег с кредитной карты;
- выявление мошеннических заявок на получение займов или кредитов;
- выявление мошеннических требований о выплате страховых возмещений;
- выявление случаев мошеннического доступа к счетам.

10.3 Самоуправляемые транспортные средства

Ожидается, что самоуправляемые, беспилотные транспортные средства в будущем могут стать обычным явлением. Сегодня многие технологии на основе искусственного интеллекта применяются в автомобилях в качестве средств помощи водителю. Ниже приведены примеры применения ИИ в транспортных средствах:

- оптимизация выбора маршрута (например, поиск наиболее быстрого маршрута с учётом текущих условий дорожного движения);
- автоматическое перестроение на другую полосу движения;
- избегание препятствий (например, автоматическое манипулирование тормозами, дроссельной заслонкой и рулевым управлением на основе интерпретации сигналов, поступающих от камер, фотоэлементов и датчиков расстояния);
- полностью автоматизированное перемещение из пункта А в пункт Б.

Автоматизированные транспортные средства полагаются на такие ИИ-технологии, как компьютерное зрение и планирование.

10.4 Прогнозная техническая поддержка

В отличие от профилактического технического обслуживания, когда обслуживание основано на ожидаемой продолжительности срока службы компонентов (например, на средней наработке на отказ), при прогнозной технической поддержке обслуживание и замена компонентов осуществляются на основе наблюдений над их текущим поведением и/или показателями работы, а также на ожидаемом сроке службы

компонентов. Ниже приведены примеры использования ИИ для прогнозной технической поддержки:

- обнаружение пустот под железнодорожными путями (что может привести к сходу с рельс);
- обнаружение потрескавшегося или повреждённого асфальта;
- выявление выходящих из строя подшипников в электродвигателях;
- выявление аномальных колебаний мощности в системах электроснабжения.

Приложение А
(справочное)
Сопоставление жизненного цикла ИИ-системы с определением жизненного цикла ИИ-системы, данным ОЭСР

В составе «Правовых инструментов» Организации экономического сотрудничества и развития (ОЭСР - Organization for Economic Co-operation and Development, OECD) была опубликована «Рекомендация Совета по искусственному интеллекту» (Recommendation of the Council on Artificial Intelligence) [26].

Данный документ включает следующий текст:

«**СОВЕТ**

… В отношении предложении Комитета по политике в области цифровой экономики (Committee on Digital Economy Policy):

I. СОГЛАШАЕТСЯ с тем, что для целей настоящей Рекомендации приведенные ниже термины нужно понимать следующим образом:

- Жизненный цикл системы: Жизненный цикл ИИ-системы включает следующие стадии:

i) «проектирование, данные и модели», которая представляет собой контекстно-зависимую последовательность, охватывающую планирование и проектирование, сбор и обработку данных, а также построение модели;

ii) «верификация и валидация»;

iii) «развёртывание»; и

iv) «эксплуатация и мониторинг».

Эти стадии часто выполняются итеративно и не обязательно последовательно. Решение о выводе ИИ-системы из эксплуатации может

быть принято в любой момент в течение стадии эксплуатации и мониторинга».

а также:

«1.4. **Робастность, безопасность и защищённость**

a) ИИ-системы должны быть надежными, безопасными и защищёнными на протяжении всего своего жизненного цикла, с тем, чтобы как в условиях нормального использования, так и предсказуемого корректного или некорректного использования, или при иных неблагоприятных условиях, - они функционировали надлежащим образом и не представляли собой неоправданно большую угрозу безопасности;

b) С этой целью организации и лица, играющие активную роль в жизненном цикле ИИ-системы, должны обеспечить отслеживаемость, в том числе в отношении наборов данных, процессов и решений, принятых на протяжении жизненного цикла ИИ-системы, - с тем, чтобы обеспечить возможность проведения уместного в конкретном контексте и соответствующего современным возможностям анализа результатов ИИ-системы и её ответов на запросы;

c) Организации и лица, играющие активную роль в жизненном цикле ИИ-системы должны, основываясь на их ролях, контексте и их способности действовать, на каждой стадии жизненного цикла ИИ-системы на постоянной основе применять систематический подход к управлению рисками с целью реагирования на связанными с ИИ-системами риски, включая риски для неприкосновенности частной жизни (персональных данных), информационной безопасности, защищённости и объективности».

На рисунке А.1 показано, как это определение жизненного цикла ИИ-системы может быть сопоставлено с жизненным циклом ИИ-системы, описанном в разделе 6:

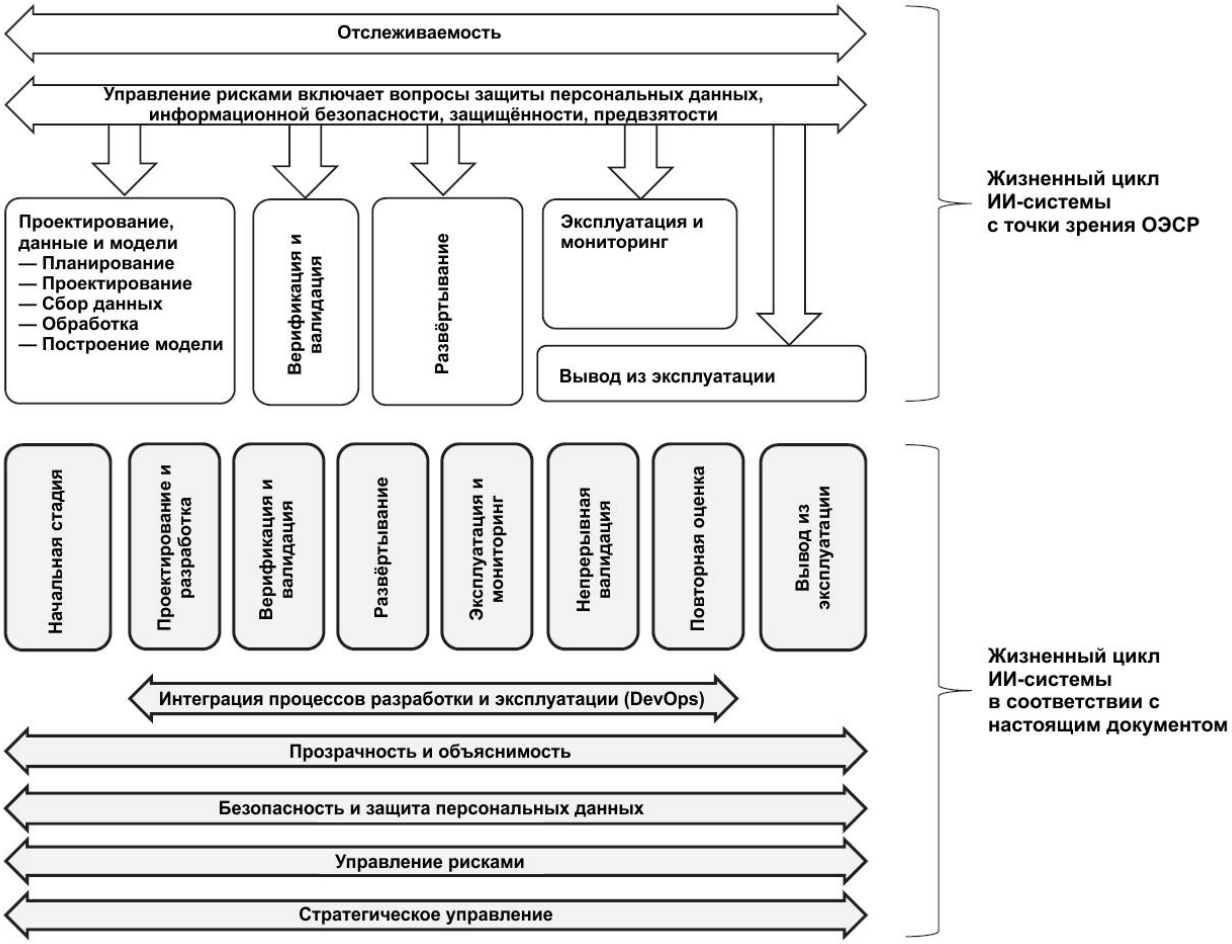

Рисунок А.1 — Сопоставление с жизненным циклом ИИ-системы согласно ОЭСР

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

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ISO/IEC 2382:2015, Information technology – Vocabulary	MOD*	ГОСТ 33707-2016 (ISO/IEC 2382:2015) «Информационные технологии. Словарь»
ISO 5127:2017, In-formation and documentation - Foundation and vocabulary	MOD**	ГОСТ 7.0-99 «Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения»
ISO/IEC/IEEE 12207:2017, Systems and software engineering - Software life cycle processes	IDT**	ГОСТ Р ИСО/МЭК 12207-2010 «Информационная технология. Системная и программная инженерия. Процессы жизненного цикла программных средств»
ISO/IEC/IEEE 15288:2023, Systems and software engineering - System life cycle processes	NEQ**	ГОСТ Р 57193-2016 / ISO/IEC/IEEE 15288:2015 «Системная и программная инженерия. Процессы жизненного цикла систем»

Продолжение таблицы ДА.1

ISO/IEC/IEEE 15289:2019, Systems and software engineering - Content of life-cycle information items (documentation)	IDT**	ГОСТ Р 58609-2019 / ISO/IEC/IEEE 15289:2017 «Системная и программная инженерия. Состав и содержание информационных элементов жизненного цикла (документации)»
ISO/IEC 17788:2014, Information technology - Cloud computing - Overview and vocabulary	IDT	ГОСТ ISO/IEC 17788-2016 «Информационные технологии. Облачные вычисления. Общие положения и терминология»
ISO/IEC 20546:2019, Information technology - Big data - Overview and vocabulary	IDT	ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии. Большие данные. Обзор и словарь»
ISO/IEC 23894:2023, Information technology - Artificial intelligence - Guidance on risk management	NEQ**	ПНСТ 776-2022 (ISO/IEC FDIS 23894) «Информационные технологии. Интеллект искусственный. Управление рисками»
ISO/IEC TR 24029-1:202, Artificial Intelligence (AI) - Assessment of the robustness of neural networks - Part 1: Overview	IDT	ГОСТ Р 70462.1-2022 / ISO/IEC TR 24029-1-2021 «Информационные технологии. Интеллект искусственный. Оценка робастности нейронных сетей. Часть 1. Обзор»
ISO/IEC 27000:2018, Information technology - Security techniques - Information security management systems - Overview and vocabulary	IDT	ГОСТ Р ИСО/МЭК 27000-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология»

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ISO/IEC 29100, Information technology - Security techniques - Privacy framework	IDT	ГОСТ ISO/IEC 29100-2021 «Информационные технологии. Методы и средства обеспечения безопасности. Основы защиты персональных данных»
ISO/IEC 30141:2018 + Cor. 1:2020, Internet of Things (IoT) - Reference Architecture	MOD	ПНСТ 438-2020 (ИСО/МЭК 30141:2018) «Информационные технологии. Интернет вещей. Типовая архитектура»,
ISO 31000:2018, Risk management – Guidelines	IDT	ГОСТ Р ИСО 31000-2019 «Менеджмент риска. Принципы и руководство»
ISO/IEC 38500, Information technology - Governance of IT for the organization	IDT	ГОСТ Р ИСО/МЭК 38500-2017 «Информационные технологии. Стратегическое управление ИТ в организации»

* В настоящей таблице использовано следующее условное обозначение степени соответствия стандарта:

NEQ — неэквивалентный стандарт

MOD — модифицированный стандарт

IDT — идентичный стандарт

** В России принят национальный стандарт, основанный на ранней редакции международного стандарта

Библиография

[1] ISO/IEC 12207:2008, *Systems and software engineering — Software life cycle processes*

[2] ISO/IEC 15288:2015, *Systems and software engineering — System life cycle processes*

[3] ISO/IEC/IEEE 15289:2019, *Systems and software engineering — Content of life-cycle information items (documentation)*

[4] ISO/IEC 17788:2014, *Information technology — Cloud computing — Overview and vocabulary*

[5] ISO/IEC 17789:2014, *Information technology — Cloud computing — Reference architecture*

[6] ISO/IEC 20546:2019, *Information technology — Big data — Overview and vocabulary*

[7] ISO/IEC 20547-3:2020, *Information technology — Big data reference architecture — Part 3: Reference architecture*

[8] ISO/IEC 20889:2018, *Privacy enhancing data de-identification terminology and classification of techniques*

[9] ISO/IEC 20924:2021, *Information technology — Internet of Things (IoT) — Vocabulary*

[10] ISO/IEC 23053, *Information technology — Artificial Intelligence (AI) — Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

[11] ISO/IEC/TR 23188:2020, *Information technology — Cloud computing — Edge computing landscape*

[12] ISO/IEC 23894, *Information technology — Artificial intelligence — Risk management*

[13] ISO/IEC/TR 24027:2021, *Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*

[14] ISO/IEC/TR 24028:2020, *Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*

[15] ISO/IEC/TR 24029-1:2021, *Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*

[16] ISO/IEC 27040:2015, *Information technology — Security techniques — Storage security*

[17] ISO/IEC 30141:2018, *Internet of Things (IoT) — Reference Architecture*

[18] Graves A., Abdel-rahman Mohamed, Geoffrey E. Hinton, Speech recognition with deep recurrent neural networks, IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, DOI: 10 .1109/ ICASSP .2013 .6638947.

[19] Jena A., Reasoners and rule engines: Jena inference support, https://jena.apache.org/documentation/inference/index.html.

[20] Artificial Intelligence Methodologies and Their Application to Diabetes. https://pubmed.ncbi.nlm.nih.gov/28539087/.

[21] Elman Jeffrey L.“, Finding structure in time.” Cognitive science **14**.2 (1990): 179-211.

[22] Hochreiter Sepp, Schmidhuber Juergen“, Long short-term memory.” Neural computation **9**.8 (1997): 1735-1780.

[23] Japanese Society of Artificial Intelligence, AI Map Beta, https://www.ai-gakkai.or.jp/pdf/aimap/AIMapEN20190606.pdf.

[24] Zadeh L.A.“, Soft computing and fuzzy logic,” IEEE Software, 1994, vol. **11**, issue 6.

[25] Rigla M., Gema García-Sáez B., Pons, M., Artificial Intelligence Methodologies and Their Application to Diabetes Hernando, Journal of diabetes science and technology, 2018, DOI: 10.1177/1932296817710475.

[26] Recommendation of the councile on artificial intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[27] Rozenblit J.W., Cognitive computing: Principles, architectures, and applications. In: Proc. 19th European Conf. on Modelling and Simulation (ECMS) (2005).

[28] S. C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); **5**:13— 22.

[29] Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach (3rd Edition) (Essex, England: Pearson, 2009).

[30] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE — On-Road Automated Driving (ORAD) committee, https://saemobilus.sae.org/content/J3016 201806/.

[31] ISO/IEC/IEEE 24765:2017, *Systems and software engineering — Vocabulary*

[32] ISO/IEC 2382:2015, *Information technology — Vocabulary*

[33] ISO 16439:2014, *Information and documentation — Methods and procedures for assessing the impact of libraries*

[34] ISO/IEC 2382-28:1995, *Information technology — Vocabulary — Part 28: Artificial intelligence — Basic concepts and expert systems*

[35] ISO 8373:2012, *Robots and robotic devices — Vocabulary*

[36] ISO 20252:2019, *Market, opinion and social research, including insights and data analytics — Vocabulary and service requirements*

[37] ISO/IEC 29100:2011/Amd1: 2018, *Information technology — Security techniques — Privacy framework — Amendment 1: Clarifications*

[38] ISO/IEC 38500:2015, *Information technology — Governance of IT for the organization*

[39] ISO/IEC 27000:2018, *Information technology — Security techniques — Information security management systems — Overview and vocabulary*

[40] ISO 31000:2018, *Risk management — Guidelines*

[41] ISO/IEC 27042:2015, *Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence*

[42] ISO 17100:2015, *Translation services — Requirements for translation services*

[43] ISO/IEC 15944-8:2012, *Information technology — Business operational view — Part 8: Identification of privacy protection requirements as external constraints on business transactions*

[44] ISO 5127:2017, *Information and documentation — Foundation and vocabulary*

[45] ISO/IEC 20071-11, *Information technology — User interface component accessibility — Part 11: Guidance on text alternatives for images*

УДК 004.8:006.354 ОКС 35.020; 01.040.35

Ключевые слова: информационные технологии (ИТ), искусственный интеллект (ИИ), большие данные, аналитика данных, терминология, ИИ-системы, жизненный цикл ИИ-систем

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

Исполнитель А. А. Храмцовская