Национальный институт сертифицированных консультантов по управлению

19/10/2026

Этический кодекс НИСКУ по взаимодействию консультантов по управлению с искусственным интеллектом

Национальный институт сертифицированных консультантов по управлению

СОДЕРЖАНИЕ

1. Общие положения *3*

1.1. Цель и сфера применения кодекса НИСКУ **3**

1.2. Термины и определения **4**

1.3. Ответственность члена НИСКУ и клиента при взаимодействии с ИИ-системами **5**

2. Взаимосвязь с Кодексом профессионального поведеня (Этическим кодексом) сертифицированного консультанта по управлению (СМС) *6*

2.1. Главенство Кодекса профессионального поведения (Этического кодекса) сертифицированного консультанта по управлению (СМС) **6**

2.2. Недопустимость освобождения от ответственности за использование ИИ **6**

2.3. Ненадлежащее использования ИИ как нарушение этических норм **7**

3. Этические принципы *7*

3.1. Профессионализм и компетентность **7**

3.2. Справедливость и борьба с дискриминацией **8**

3.3. Честность, объективность и прозрачность **8**

3.4. Конфиденциальность и уважение прав человека **8**

3.5. Безопасность и управление рисками **9**

3.6. Отчетность и документация **10**

3.7. Социальная ответственность и благополучие **10**

4. Взаимодействие с ИИ *10*

4.1. Определение целей и границ применения ИИ **10**

4.2. Выбор подходящих ИИ-технологий и алгоритмов **11**

4.3. Качество данных для обучения и валидации ИИ-систем **13**

4.4. Тестирование и аудит ИИ-решений на надежность и безопасность **14**

4.5. Прозрачность ИИ и возможность объяснения логики его работы **16**

4.6. Контроль за принятием решений ИИ и подотчетность **17**

4.7. Мониторинг ИИ-систем в процессе эксплуатации **18**

4.8. Обеспечение функционирования изолированной системы ИИ (закрытого контура) **20**

5. Работа с заказчиком *20*

Национальный институт сертифицированных консультантов по управлению

5.1. Информирование клиента о возможностях и ограничениях ИИ 21

5.2. Определение задач и метрик эффективности ИИ-решений 23

5.3. Учет рисков и обязательств, связанных с внедрением ИИ 23

5.4. Обеспечение конфиденциальности и безопасности данных клиента 25

5.5. Передача знаний и обучение команды заказчика.....26

6. Профессиональное развитие членов НИСКУ.....26

6.1. Повышение уровня экспертизы в области ИИ-технологий.....27

6.2. Развитие этических компетенций 27

6.3. Обмен знаниями и лучшими практиками в профессиональном сообществе.....28

Приложения 28

Приложение 1 Рекомендуемый Чек-лист оценки соответствия ИИ-решения требованиям EU AI Act 28

Национальный институт сертифицированных консультантов по управлению

1. ОБЩИЕ ПОЛОЖЕНИЯ

1.1. ЦЕЛЬ И СФЕРА ПРИМЕНЕНИЯ КОДЕКСА НИСКУ

Настоящий Этический кодекс члена НИСКУ по взаимодействию с искусственным интеллектом (далее – Кодекс) носит рекомендательный характер и устанавливает основные этические принципы и нормы поведения, которыми должны руководствоваться члены НИСКУ при оказании услуг, связанных со внедрением и использованием систем искусственного интеллекта (ИИ) в организациях.

Кодекс основан на принципах и нормах поведения консультантов по управлению, сформулированных в Кодексе профессионального поведения (Этическом кодексе) членов НИСКУ, а также на основных положениях Кодекса ответственного взаимодействия консультантов по управлению с ИИ, разработанного ICMCI[1].

Цель Кодекса – обеспечить ответственное и прозрачное применение ИИ-технологий, защиту интересов клиентов, общества и окружающей среды, а также содействовать устойчивому развитию и инновациям.

Положения данного кодекса учитывает необходимость соблюдения правовых норм, непосредственно регулирующих соответствующие вопросы, в особенности правовых норм, содержащихся в следующих нормативных правовых актах: Федеральный закон «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных» от 24.04.2020 № 123-ФЗ, Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ, Федеральный закон «О персональных данных» от 27.07.2006 № 152-ФЗ, Федеральный закон от 26.07.2026 № 243-ФЗ «О поддержке развития технологий искусственного интеллекта в Российской Федерации».

Положения данного Кодекса применимы ко всем членам НИСКУ, которые в своей профессиональной деятельности взаимодействуют с ИИ-системами, независимо от их специализации, должности и организационно-правовой формы.

Кодекс распространяется на отношения, связанные с этическими аспектами создания (проектирования, конструирования, пилотирования), внедрения и использования технологий ИИ на всех этапах жизненного цикла, которые в настоящее время не урегулированы законодательством Российской Федерации и/или актами технического регулирования.

Все положения носят рекомендательный характер и не являются единственным правильным вариантом поведения, однако вытекают из логики этики делового поведения.

[1] ICMCI Code of Responsible AI Interaction for Management Consultants (доступен на официальном сайте ICMCI [Электронный ресурс], URL: https://www.icmci.org).

1.2. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем Кодексе используются следующие термины и определения:

Организация экономического сотрудничества и развития определяет систему искусственного интеллекта как “компьютерную систему, которая для достижения явных или неявных целей на основе полученных входных данных определяет, как генерировать выходные данные, такие как прогнозы, контент, рекомендации или решения, которые могут влиять на физическую или виртуальную среду. Различные системы искусственного интеллекта различаются по уровню автономии и адаптивности после развертывания”[2].

- Искусственный интеллект (ИИ) – свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека.

- ИИ-система – программное обеспечение, основанное на методах искусственного интеллекта, включая машинное обучение, глубокое обучение, обработку естественного языка, компьютерное зрение и т.д.

- Ответственный ИИ –комплексное понятие, обозначающее системный подход к разработке и внедрению искусственного интеллекта, который основывается на управлении рисками и направлен на обеспечение надежности, безопасности, прозрачности, справедливости и подотчетности ИИ-систем на протяжении всего их жизненного цикла.

- Управляемая автономия – принцип, обеспечивающий способность системы искусственного интеллекта функционировать и принимать решения с определенной степенью независимости, но в установленных рамках, определяемых политиками, правилами, этическими нормами и системами контроля для минимизации рисков

- Состязательная атака – злонамеренный метод, который манипулирует моделями машинного обучения путем намеренной подачи им обманчивых данных, чтобы вызвать некорректное или непреднамеренное поведение. Эти атаки используют уязвимости в базовой логике модели, часто за счет незаметных для человека изменений во входных данных .

- Высокорисковый ИИ – это категория систем искусственного интеллекта, которые в силу своих характеристик или предназначения создают значительный потенциал причинения вреда здоровью, безопасности или фундаментальным правам человека и поэтому подпадают под действие строгих обязательств в соответствии с Регламентом ЕС об ИИ (EU AI Act).

- Модель «белого ящика» — открытая прозрачная система с полностью известным внутренним устройством, логику и алгоритмы работы которой можно легко отследить и объяснить.

[2] OECD AI Principles overview [Электронный ресурс]. URL: https://oecd.ai/en/ai-principles

- Модель «черного ящика» — непрозрачная система со скрытой внутренней структурой, в которой известны только входные данные и конечный результат, но не сам процесс вычислений.

- Член НИСКУ – физическое или юридическое лицо, оказывающее консультационные услуги, связанные с внедрением и использованием ИИ-систем в организациях.

- Клиент – организация, которая пользуется услугами члена НИСКУ с целью внедрения ИИ-решений.

- Стейкхолдеры – все заинтересованные стороны, на которых прямо или косвенно влияет использование ИИ-систем, включая сотрудников, клиентов, партнеров, государственные органы, общественность.

1.3. ОТВЕТСТВЕННОСТЬ ЧЛЕНА НИСКУ И КЛИЕНТА ПРИ ВЗАИМОДЕЙСТВИИ С ИИ-СИСТЕМАМИ

Член НИСКУ отвечает за качество предоставляемых услуг, связанных с ИИ, перед клиентом и обществом, в части своего уровня ответственности. Он действует добросовестно, с должной осмотрительностью и соблюдением применимого законодательства.

Членам НИСКУ следует ответственно подходить к вопросам влияния ИИ на общество и граждан на каждом этапе его жизненного цикла, включая неприкосновенность частной жизни, этичное, безопасное и ответственное использование персональных данных, к характеру, степени и размеру ущерба, который может последовать в результате использования технологий и ИИ, а также при выборе и использовании аппаратных средств и программного обеспечения, задействованных на различных жизненных циклах.

Члены НИСКУ, в случае применения ИИ, принимают на себя обязательство следовать процессу уведомления регулятора и заказчика об инцидентах с ИИ в соответствии с законодательством и Российской Федерации и подзаконными актами федеральных органов государственной власти, в том числе разъяснительных актов Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций.

Членам НИСКУ рекомендуется учитывать наличие механизма эскалации рисков ИИ «Цепочка ответственности»:

1. Выявление: консультант фиксирует риск и немедленно уведомляет своего непосредственного руководителя или заказчика.

2. Анализ: руководитель/заказчик анализирует риск. Риски уровня «Высокий»/«Критический» в течение 48 часов и направляет его ответственному лицу.

3. Решение: ответственное лицо оценивает риск и определяет меры по его устранению. Критические случаи эскалирует руководителю/заказчику.

4. Стратегия: руководитель/заказчик принимает финальное решение и обеспечивает его исполнение.

Клиент, в свою очередь, принимает на себя ответственность за определение целей и задач внедрения ИИ, предоставление необходимых данных и ресурсов, а также за принятие решений на основе рекомендаций члена НИСКУ. Клиент должен обеспечить прозрачность использования ИИ для сотрудников и клиентов.

Распределение ответственности между членом НИСКУ и клиентом, а также другими третьими сторонами, должно быть четко определено в соглашении об оказании консультационных услуг с учетом требований данного Кодекса.

2. ВЗАИМОСВЯЗЬ С КОДЕКСОМ ПРОФЕССИОНАЛЬНОГО ПОВЕДЕНИЯ (ЭТИЧЕСКИМ КОДЕКСОМ) СЕРТИФИЦИРОВАННОГО КОНСУЛЬТАНТА ПО УПРАВЛЕНИЮ (CMC)

2.1. ГЛАВЕНСТВО КОДЕКСА ПРОФЕССИОНАЛЬНОГО ПОВЕДЕНИЯ (ЭТИЧЕСКОГО КОДЕКСА) СЕРТИФИЦИРОВАННОГО КОНСУЛЬТАНТА ПО УПРАВЛЕНИЮ (CMC)

Настоящий Кодекс основан на Кодексе профессионального поведения (Этическом кодексе) сертифицированного консультанта по управлению (CMC)[3] (далее – Генеральный кодекс) и дополняет его. Он разъясняет, каким образом установленные обязанности Кодекса применяются при использовании систем искусственного интеллекта в профессиональной консультационной деятельности.

Все обязательства, вытекающие из Генерального кодекса, включая основные обязанности соблюдения честности, объективности, компетентности, конфиденциальности, а также ответственности перед клиентами, обществом и профессией, продолжают действовать в полном объеме при использовании ИИ.

Использование ИИ в структурах НИСКУ и сертифицированными управленческими консультантами (CMC) также регулируется соответствующими системами, предназначенными для управления рисками в НИСКУ. Риски, связанные с ИИ, должны выявляться, оцениваться и управляться в соответствии с допустимым уровнем риска организации и установленными процессами управления.

2.2. НЕДОПУСТИМОСТЬ ОСВОБОЖДЕНИЯ ОТ ОТВЕТСТВЕННОСТИ ЗА ИСПОЛЬЗОВАНИЕ ИИ

[3] ICMCI Code of Responsible AI Interaction for Management Consultants (доступен на официальном сайте ICMCI [Электронный ресурс], URL: https://www.icmci.org).

Использование системы ИИ не ограничивает, не смягчает и не освобождает от личной и профессиональной ответственности консультанта или фирмы. Консультанты несут полную ответственность за результаты работы, рекомендации и их последствия, независимо от того, использовались ли инструменты ИИ при их подготовке.

Если система ИИ вносит вклад в конечный результат, консультант, который представляет, передает или полагается на этот результат, несет ту же профессиональную ответственность, как если бы работа была выполнена исключительно человеком.

2.3. НЕНАДЛЕЖАЩЕЕ ИСПОЛЬЗОВАНИЕ ИИ КАК НАРУШЕНИЕ ЭТИЧЕСКИХ НОРМ

Ненадлежащее использование ИИ в контексте профессиональной консультационной деятельности рассматривается как нарушение Генерального кодекса, аналогичное традиционным формам профессионального проступка. Действия, которые могут квалифицироваться как ненадлежащее использование ИИ, включают, но не ограничиваются:

- представление контента, сгенерированного ИИ, как собственной работы консультанта без надлежащей проверки, адаптации и, при необходимости, указания источника;

- представление непроверенных или по существу неподтвержденных результатов работы ИИ в качестве готового профессионального продукта;

- использование ИИ для обхода требований к компетенции, обучению или оценке, включая требования для получения статуса CMC;

- ввод конфиденциальных данных клиентов в системы ИИ в нарушение обязательств по защите данных или инструкций клиента; и использование систем ИИ способом, который с достаточной очевидностью ведет к дискриминации отдельных лиц или групп или причинению им вреда.

3. ЭТИЧЕСКИЕ ПРИНЦИПЫ

3.1. ПРОФЕССИОНАЛИЗМ И КОМПЕТЕНТНОСТЬ

Член НИСКУ обладает необходимыми знаниями, навыками и опытом в области ИИ-технологий, чтобы предоставлять качественные услуги и рекомендации клиентам. Он постоянно повышает свою квалификацию, следит за развитием ИИ и лучшими практиками в индустрии.

Члены НИСКУ по своему желанию могут получить сертификации по стандарту IEEE P7000-серии (выявление возможных этических проблем при разработке и внедрении

высокотехнологичных продуктов) и ISO 38507 (управление применением ИИ в организации), а также участвовать в регулярном аудите компетенций в области применения ИИ.

Член НИСКУ трезво оценивает свои возможности и ограничения, берется только за те проекты, которые соответствуют его компетенциям. В случае недостатка экспертизы он привлекает более опытных коллег или отказывается от задания.

3.2. СПРАВЕДЛИВОСТЬ И БОРЬБА С ДИСКРИМИНАЦИЕЙ

ИИ-модели могут воспроизводить предвзятость и дискриминацию, заложенную в исторических данных, по признаку пола, расы, возраста, происхождения и других характеристик. Член НИСКУ прилагает усилия для выявления и устранения такой предвзятости.

При разработке и аудите ИИ-решений член НИСКУ использует репрезентативные и сбалансированные данные, тестирует модели на дискриминацию, применяет методы обеспечения равенства и многообразия. ИИ должен быть справедливым и беспристрастным.

Члены НИСКУ вправе проводить регулярный внешний аудит на предмет выявления смещений в системах ИИ, соответствующего критериям стандарта IEEE P7003 по выявлению возможностей возникновения этических проблем.

3.3. ЧЕСТНОСТЬ, ОБЪЕКТИВНОСТЬ И ПРОЗРАЧНОСТЬ

Член НИСКУ действует честно и объективно, предоставляет клиентам полную и достоверную информацию об ИИ-решениях, включая их возможности, ограничения и потенциальные риски. Он не вводит заказчика в заблуждение, не дает необоснованных обещаний и не скрывает важные факты.

Методы и результаты работы члена НИСКУ прозрачны и проверяемы. Он документирует свои действия, решения и рекомендации, а также предоставляет клиенту доступ к исходному коду, моделям и данным в той мере, в которой это не нарушает права интеллектуальной собственности и конфиденциальности.

3.4. КОНФИДЕНЦИАЛЬНОСТЬ И УВАЖЕНИЕ ПРАВ ЧЕЛОВЕКА

Член НИСКУ уважает конфиденциальность информации, полученной от клиента в ходе оказания услуг, и не разглашает ее третьим лицам без явного согласия заказчика, кроме случаев, предусмотренных законом.

Член НИСКУ может использовать конфиденциальность по умолчанию — подход, при котором защита данных встраивается в архитектуру системы и бизнес-процессы на этапе проектирования, а не добавляется постфактум.

При работе с данными, особенно содержащими персональную информацию, член НИСКУ соблюдает принципы минимизации, целевого использования, ограничения хранения и обеспечения безопасности данных. Он помогает клиенту выполнять требования законодательства о защите данных.

ИИ-системы не должны нарушать фундаментальные права человека, включая право на приватность, свободу слова, равенство перед законом и др. Член НИСКУ анализирует разрабатываемые решения на предмет соблюдения прав человека.

Член НИСКУ действует с должной осмотрительностью (совершает комплекс проверочных и аналитических действий в отношении субъектов и объектов отношений, а именно превентивно оценивает, контролирует и минимизирует риски нарушения прав человека со стороны внедряемых ИИ-систем на всех этапах их жизненного цикла). При этом член НИСКУ руководствуется положениями международного стандарта ISO 26000:2010 «Руководство по социальной ответственности».

В случае, если ИИ поспособствует нарушению прав человека, например, при использовании систем распознавания, член НИСКУ предупреждает об этом клиента и предлагает альтернативы и механизмы защиты прав.

3.5. БЕЗОПАСНОСТЬ И УПРАВЛЕНИЕ РИСКАМИ

Член НИСКУ идентифицирует и оценивает риски, связанные с ИИ-системами, включая риски причинения вреда, нарушения прав, финансовых потерь, репутационного ущерба и т.д. Он рекомендует клиенту меры по управлению этими рисками.

Члены НИСКУ обеспечивают максимально возможную защиту от несанкционированного вмешательства в работу ИИ третьих лиц; внедряют адекватные технологии информационной безопасности, в том числе применяют внутренние механизмы защиты ИИ от несанкционированных вмешательств и информирования пользователей и разработчиков о таких вмешательствах; содействуют информированию пользователей о правилах информационной безопасности при использовании ИИ.

ИИ-решения должны быть надежными и безопасными на протяжении всего жизненного цикла. Член НИСКУ следует лучшим практикам безопасной разработки, тестирования системы на устойчивость к взлому и атакам, внедрения механизмы контроля и аварийного отключения ИИ с участием ИТ подразделений заказчика.

Для критических приложений ИИ-систем в особых случаях приветствуется проведение оценки рисков посредством привлечения нейтральной третьей стороны или уполномоченного официального органа, но без ущерба для работоспособности и информационной безопасности такой ИИ-системы, а также охраны интеллектуальной собственности и коммерческой тайны разработчика.

3.6. ОТЧЕТНОСТЬ И ДОКУМЕНТАЦИЯ

Организации в лице ответственных консультантов документируют важные решения, принятые с использованием ИИ, а также обоснование выбора конкретных инструментов и результатов их работы – в объеме, соразмерном рискам и значимости выполняемой задачи.

Организации устанавливают четкое разграничение ответственности за управление использованием ИИ, соответствующее масштабу и сложности деятельности организации.

Отдельные консультанты несут ответственность за то, чтобы использование ими ИИ соответствовало настоящему Кодексу и внутренним правилам их фирмы.

3.7. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ И БЛАГОПОЛУЧИЕ

Член НИСКУ учитывает потенциальное влияние ИИ-систем на благополучие людей и общества в целом. Он учитывает экологическое, социальное и экономическое воздействие выбираемых инструментов и систем, а также содействует их соразмерному и ответственному использованию в соответствии с профессиональными и гражданскими обязательствами. Он стремится к тому, чтобы ИИ приносил пользу и не причинял вреда, особенно уязвимым слоям населения.

Член НИСКУ при наличии оснований вправе проводить общественные консультации для проектов с высоким социальным риском (High-risk AI).

Член НИСКУ предупреждает клиента о возможных негативных последствиях использования ИИ, таких как потеря рабочих мест, дискриминация, усиление неравенства и т.д. Он рекомендует меры по их предотвращению или смягчению.

Члены НИСКУ не допускают передачи полномочий ответственного нравственного выбора ИИ-системам, не делегируют им ответственность за последствия принятия решений – за все последствия работы с ИИ всегда отвечает человек (физическое или юридическое лицо, признаваемое субъектом ответственности в соответствии с действующим законодательством Российской Федерации). Членам НИСКУ рекомендуется принимать все меры для определения ответственности конкретных участников жизненного цикла ИИ-систем с учетом их роли и специфики каждого этапа.

4. ВЗАИМОДЕЙСТВИЕ С ИИ

4.1. ОПРЕДЕЛЕНИЕ ЦЕЛЕЙ И ГРАНИЦ ПРИМЕНЕНИЯ ИИ

Для четкого определения бизнес-целей и желаемых результатов, которые предполагается достичь с помощью внедрения ИИ-технологий, членом НИСКУ инициируется диалог с клиентом. Этот процесс должен включать:

a) Анализ текущего состояния бизнес-процессов, ИТ-инфраструктуры и уровня цифровой зрелости организации клиента;

b) Идентификацию проблемных областей, узких мест и возможностей для улучшения, которые могут быть адресованы с применением ИИ;

c) Определение ключевых показателей эффективности (КПЭ) и критериев успеха для оценки достижения поставленных целей.

Член НИСКУ предоставляет клиенту исчерпывающую информацию о возможностях и ограничениях современных ИИ-технологий применительно к его отрасли и специфике деятельности. Он проводит всесторонний анализ условий, при которых использование ИИ является целесообразным, обоснованным и отвечает этическим требованиям, учитывая такие факторы как:

a) Наличие и качество исторических данных, необходимых для обучения ИИ-моделей;

b) Уровень алгоритмизируемости и формализуемости решаемых задач;

c) Потенциальное влияние ИИ на бизнес-процессы, рабочие места и взаимодействие с клиентами;

d) Риски и последствия ошибочных решений ИИ для жизни и здоровья людей, окружающей среды, финансового состояния и репутации организации;

e) Нормативно-правовые требования и ограничения, связанные с использованием ИИ в конкретных сферах;

f) Уровень доверия и готовности сотрудников и клиентов к взаимодействию с ИИ-системами.

Если применение ИИ признано нецелесообразным или преждевременным, член НИСКУ рекомендует клиенту альтернативные подходы к решению поставленных задач, такие как оптимизация процессов, автоматизация рутинных операций, традиционные методы анализа данных и поддержки принятия решений.

Члены НИСКУ осуществляют добросовестное информирование пользователей об их взаимодействии с ИИ-системами, когда это затрагивает вопросы прав человека и критических сфер его жизни, и обеспечивать возможность прекратить такое взаимодействие по желанию пользователя.

4.2. ВЫБОР ПОДХОДЯЩИХ ИИ-ТЕХНОЛОГИЙ И АЛГОРИТМОВ

Член НИСКУ обладает экспертными знаниями и практическим опытом в области современных ИИ-технологий, включая машинное обучение, глубокие нейронные сети, обработку естественного языка, компьютерное зрение и др. Он непрерывно следит за развитием индустрии ИИ, научными публикациями и лучшими практиками.

При выборе конкретных ИИ-методов и инструментов для решения бизнес-задач клиента член НИСКУ руководствуется следующими принципами:

a) Соответствие ИИ-технологий характеру и сложности решаемых задач, объему и структуре доступных данных;

b) Обоснованность и доказанная эффективность выбранных алгоритмов и архитектур ИИ для данной предметной области;

c) Прозрачность и интерпретируемость ИИ-моделей, обеспечивающие возможность объяснения логики их работы и принимаемых решений (использование моделей «белого ящика» - прозрачных систем с полностью известным внутренним устройством, логику и алгоритмы работы которой можно легко отследить и объяснить);

d) Воспроизводимость результатов ИИ и возможность их независимой проверки (аудита) в случае использования моделей «черного ящика» - непрозрачных систем со скрытой внутренней структурой, в которой известны только входные данные и конечный результат, но не сам процесс вычислений;

e) Справедливость и отсутствие дискриминации от ИИ-моделей по отношению к различным группам населения в случае использования закрытых моделей «черного ящика»;

f) Устойчивость ИИ-систем к атакам и попыткам манипулирования со стороны злоумышленников (закрытые модели «черного ящика»);

g) Экономическая целесообразность и эффективность внедрения выбранных решений с учетом совокупной стоимости владения.

Член НИСКУ предоставляет клиенту обоснование выбора конкретных ИИ-технологий и платформ с указанием их преимуществ, ограничений и потенциальных рисков. При наличии нескольких подходящих вариантов следует организовать их сравнительное тестирование на реальных или синтетических данных клиента с оценкой ключевых показателей качества и затрат. В случае, если выбор ИИ- технологий берет на себя заказчик, член НИСКУ предусматривает в контракте возможность доступа к таким обоснованиям выбора.

В случаях, когда применение ИИ сопряжено с высокими рисками для критически важных систем, здоровья и безопасности людей или основных прав человека, член НИСКУ рекомендует максимально прозрачные и интерпретируемые открытые модели «белого ящика», такие как деревья решений, логистическая регрессия, наивный байесовский классификатор, матрицу оценки TCO и E2E-интерпретируемости (сквозное тестирование по всему процессу разработки и внедрения) и др. Если задача требует использования более сложных подходов, таких как глубокое обучение, внедряются механизмы объяснения их работы (для компенсации недостатка моделей «черного ящика»).

Такая деятельность осуществляется членом НИСКУ в случае наличия соответствующей подготовки и опыта, а также при наличии запроса от клиента.

4.3. КАЧЕСТВО ДАННЫХ ДЛЯ ОБУЧЕНИЯ И ВАЛИДАЦИИ ИИ-СИСТЕМ

Качество и репрезентативность данных, на которых обучаются и тестируются ИИ-модели, являются ключевыми факторами, определяющими их эффективность, справедливость и устойчивость к изменениям.

Член НИСКУ оказывает клиенту содействие на всех этапах работы с данными, включая:

a) Определение потребностей в данных, исходя из целей и архитектуры ИИ-решения, а также выявление имеющихся у клиента источников релевантной информации;

b) Сбор, интеграцию и структурирование данных из различных внутренних и внешних систем клиента, включая корпоративные базы данных, веб-сайты, социальные сети, InternetofThings -IoT-устройства (облачные технологии) и др.;

c) Очистку и предварительную обработку данных, включая обнаружение и удаление дубликатов, противоречий, пропусков, выбросов и аномалий;

d) Обогащение данных путем извлечения дополнительных признаков, связывания с внешними источниками, генерации синтетических примеров и др.;

e) Разделение данных на обучающую, валидационную и тестовую выборки в соответствии с лучшими практиками и особенностями решаемой задачи;

f) Документирование и версионирование наборов данных, обеспечение их безопасного хранения и контроля доступа, включая их регистрацию в Data Catalog с указанием Data Lineage.

На этапе предварительного анализа член НИСКУ исследует доступные данные с помощью методов описательной статистики и визуализации, чтобы выявить их основные свойства, распределения значений, наличие корреляций и трендов. Он оценивает качество данных по таким критериям как:

a) Полнота и достаточность данных для решения поставленной задачи с требуемым уровнем точности;

b) Сбалансированность и репрезентативность данных по отношению к целевой популяции или процессу;

c) Согласованность и непротиворечивость данных во времени и между различными источниками;

d) Точность и надежность данных, отсутствие существенных ошибок и искажений;

e) Актуальность и своевременность данных для принятия решений;

f) Отсутствие переобучения и утечки информации между обучающей и тестовой выборками.

В случае выявления проблем с качеством и полнотой данных член НИСКУ предлагает клиенту варианты их решения, в том числе:

a) Целенаправленный сбор недостающих данных в рамках специальных кампаний или краудсорсинговых инициатив;

b) Синтезирование дополнительных тренировочных примеров с помощью техник аугментации, переноса обучения и генеративных состязательных сетей;

c) Использование техник обучения с частичным привлечением учителя и активного обучения для эффективной разметки данных экспертами.

Член НИСКУ может устанавливать и документировать требования к качеству данных на всех этапах жизненного цикла ИИ-системы, а также проводить независимую экспертизу в случае закрепления его ответственности в контракте. В случае ответственности клиента за качество данных, член НИСКУ четко определяет область своего участия в условиях контракта.

Такая деятельность осуществляется членом НИСКУ в случае наличия соответствующей подготовки и опыта, а также при наличии запроса от клиента.

4.4. ТЕСТИРОВАНИЕ И АУДИТ ИИ-РЕШЕНИЙ НА НАДЕЖНОСТЬ И БЕЗОПАСНОСТЬ

Перед внедрением в промышленную эксплуатацию ИИ-решения проходят через многоступенчатое тестирование, чтобы подтвердить их функциональную корректность, производительность, безопасность и соответствие этическим нормам. Член НИСКУ при наличии соответствующей подготовки может планировать и организовывать комплекс тестовых активностей, охватывающий:

a) Модульное тестирование отдельных компонентов ИИ-системы (моделей, сервисов, пользовательских интерфейсов) для проверки их работы на наборах синтетических и реальных данных;

b) Интеграционное тестирование для контроля взаимодействия между компонентами ИИ-системы и ее интерфейсов с внешними ИТ-системами клиента;

c) Функциональное тестирование сценариев использования ИИ в бизнес-процессах для подтверждения соответствия поведения системы спецификациям;

d) Тестирование производительности и масштабируемости ИИ-системы при различных уровнях нагрузки, в том числе пиковой;

e) Тестирование безопасности для оценки защищенности ИИ-моделей и инфраструктуры от потенциальных угроз конфиденциальности, целостности и

доступности, таких как атаки на модель, отравление данных, инверсия модели, извлечение конфиденциальной информации;

f) Тестирование на справедливость и отсутствие дискриминации при принятии решений в отношении различных демографических групп и защищенных признаков;

g) Тестирование на устойчивость ИИ-моделей к изменениям в распределении входных данных, возникающим с течением времени или в результате внешних воздействий;

h) Тестирование на присутствие человеческих предубеждений, заложенных в данные и алгоритмы ИИ вследствие некорректного выбора признаков, недостаточно репрезентативной выборки или ошибок разметки;

i) Испытание ИИ-систем с привлечением пользователей из числа сотрудников и клиентов НИСКУ для оценки удобства использования, этичности и приемлемости их поведения.

Результаты тестирования, включая обнаруженные дефекты и несоответствия, документируются и используются для итеративного улучшения качества системы. При выявлении критических уязвимостей или этических проблем ИИ-решение не должно допускаться ко внедрению до их устранения.

Помимо предварительного тестирования, в процессе эксплуатации ИИ-систем может проводиться их регулярный аудит силами независимых экспертов. Такой аудит включает:

a) Мониторинг ключевых показателей эффективности и сравнение их фактических значений с ожидаемыми для своевременного выявления деградации качества моделей.

b) Проверку журналов и отчетов безопасности для обнаружения возможных атак и аномального поведения системы.

c) Опросы пользователей и сбор обратной связи об их удовлетворенности, доверии и проблемах при работе с ИИ.

d) Объяснение и интерпретацию принимаемых ИИ решений для оценки их обоснованности и справедливости.

e) Стресс-тестирование ИИ-системы на граничные и нетипичные случаи для контроля устойчивости.

f) Этическую экспертизу долгосрочных эффектов от использования ИИ для общества и окружающей среды.

Результаты аудита предоставляются руководству или заказчику и служат основанием для принятия решений об изменении параметров моделей, переобучении на новых данных, изменении сценариев использования или полном выводе ИИ-системы из эксплуатации. В критических случаях может потребоваться экстренная приостановка работы ИИ для предотвращения потенциального ущерба.

Такая деятельность осуществляется членом НИСКУ в случае наличия соответствующей подготовки и опыта, а также при наличии запроса от клиента.

4.5. ПРОЗРАЧНОСТЬ ИИ И ВОЗМОЖНОСТЬ ОБЪЯСНЕНИЯ ЛОГИКИ ЕГО РАБОТЫ

Обеспечение прозрачности и возможности интерпретации решений, принимаемых ИИ-системами, является ключевым фактором доверия к ним со стороны пользователей и регуляторов. Член НИСКУ прикладывает максимум усилий, чтобы ИИ-решения, разрабатываемые и внедряемые для клиента, соответствовали принципам объяснимого ИИ (XAI - eXplainable AI). Для этого необходимо:

a) Использовать по возможности простые и интуитивно понятные архитектуры ИИ-моделей, такие как деревья решений, правила, линейные модели, вероятностные графические модели, которые обладают встроенной интерпретируемостью;

c) Разрабатывать пользовательские интерфейсы, позволяющие человеку понять, на основании каких факторов ИИ пришел к тому или иному выводу, и какова степень уверенности модели в своем прогнозе;

d) Предоставлять возможность отслеживания пути принятия решений ИИ для каждого конкретного случая (например, почему заявка на кредит была отклонена) с указанием использованных данных и правил вывода;

e) Документировать все допущения, ограничения и потенциальные ошибки моделей ИИ, а также сценарии, в которых их применение может привести к нежелательным последствиям.

Уровень детализации объяснений и целевая аудитория определяются в зависимости от критичности задачи и требований регуляторов. В областях с высокими рисками, таких как медицинская диагностика, выдача кредитов, оценка кандидатов при найме, решения ИИ должны быть понятны экспертам в предметной области и поддаваться аудиту. В менее критичных сферах может быть достаточно высокоуровневых объяснений для конечных пользователей.

Если применение методов оценки невозможно или существенно снижает качество моделей, консультант организовывает альтернативные механизмы контроля и подотчетности ИИ-решений, такие как:

a) Экспертиза используемых наборов данных, признаков и алгоритмов обучения на предмет полноты, репрезентативности и отсутствия дискриминации;

b) Многократное тестирование моделей на различных подвыборках реальных и синтетических данных для проверки устойчивости и непредвзятости их работы;

c) Установление человеческого контроля в точках принятия критических решений с возможностью корректировки результатов ИИ;

d) Реализация функции отката решений ИИ и компенсации их потенциальных негативных последствий для пользователей.

Такая деятельность осуществляется членом НИСКУ в случае наличия соответствующей подготовки и опыта, а также при наличии запроса от клиента.

Член НИСКУ проводит регулярное обучение сотрудников клиента, отвечающих за эксплуатацию и поддержку ИИ-систем, методам интерпретации их поведения и устранения проблем.

4.6. КОНТРОЛЬ ЗА ПРИНЯТИЕМ РЕШЕНИЙ ИИ И ПОДОТЧЕТНОСТЬ

Несмотря на растущие возможности ИИ-систем выполнять сложные когнитивные задачи, сопоставимые с человеческими, окончательная ответственность за последствия их решений должна оставаться на людях и организациях, разрабатывающих и использующих эти системы. Член НИСКУ в этих обстоятельствах помогает клиенту выстроить эффективную систему управления ИИ, обеспечивающую баланс между автономностью интеллектуальных агентов и контролем со стороны человека. Эта система включает:

a) Четкое распределение ролей и обязанностей между внутренними подразделениями клиента (ИТ, бизнес-заказчики, службы риск-менеджмента, комплаенса, внутреннего аудита) и внешними поставщиками ИИ-решений на всех этапах их жизненного цикла;

b) Формализованные политики и процедуры разработки, тестирования, приемки, эксплуатации, мониторинга и вывода из эксплуатации ИИ-систем, учитывающие специфику предметных областей и регуляторные требования;

c) Количественные и качественные критерии оценки эффективности и рисков ИИ-решений, пороговые значения метрик, при которых требуется вмешательство человека или приостановка работы системы;

d) Матрицу принятия решений, определяющую уровни автоматизации и необходимость человеческого подтверждения в зависимости от типа решений, их влияния и затрагиваемых стейкхолдеров;

e) Механизмы аудита и протоколирования работы ИИ-систем, позволяющие восстановить картину событий при возникновении инцидентов или спорных ситуаций;

f) Планы обеспечения непрерывности бизнеса, восстановления после сбоев и реагирования на инциденты ИИ, включая оповещение клиентов, регуляторов и общественности.

Степень контроля за ИИ-системами должна соответствовать уровню потенциальных рисков и требованиям соответствующих стандартов и практик. Так, для приложений ИИ в регулируемых отраслях, таких как здравоохранение или финансы, может потребоваться валидация моделей со стороны уполномоченных органов перед допуском в эксплуатацию.

Там, где решения ИИ могут существенно повлиять на жизнь и благополучие людей, например, при постановке медицинских диагнозов или выдаче социальных льгот, они дополняются механизмами оспаривания и адвокации для пользователей (human-in-the-loop). Human-in-the-loop представляет подход, когда человек является участником процесса принятия решений, проверяет ИИ-решения в процессе работы автоматизированных систем. После этого ИИ продолжает работу. Пользователи, подвергшиеся негативному воздействию решений ИИ, должны иметь возможность запросить дополнительную информацию, обратиться к человеку-оператору или подать апелляцию.

Для задач, не требующих одобрения каждого решения человеком, определяется владелец ИИ-системы из числа высококвалифицированных специалистов клиента, обладающих как техническими знаниями, так и пониманием бизнес-контекста и способный интерпретировать поведение модели и вмешаться в ее работу при необходимости (human-on-the-loop). Человек контролирует ситуацию не на каждом этапе рабочего процесса, а при возникновении необходимости.

Член НИСКУ при необходимости обеспечивает полную прозрачность своих взаимоотношений с клиентом в рамках проектов по внедрению ИИ, включая раскрытие информации о бизнес-модели, связях с поставщиками решений, распределении прав интеллектуальной собственности, и осуществляет деятельность в соответствии с этическими нормами своей профессии.

В случае возникновения неблагоприятных последствий в результате использования ИИ-решений, разработанных при участии члена НИСКУ, он незамедлительно уведомляет клиента и предпринимает все зависящие от него меры по минимизации ущерба и предотвращению повторного возникновения подобных ситуаций, не дожидаясь окончания формальных разбирательств.

4.7. МОНИТОРИНГ ИИ-СИСТЕМ В ПРОЦЕССЕ ЭКСПЛУАТАЦИИ

Вывод ИИ-решений в продуктивную эксплуатацию не означает завершение работы по обеспечению их качества и соответствия ожиданиям бизнеса. ИИ-модели, основанные на машинном обучении, могут со временем деградировать из-за изменений в распределении входных данных или внутренней структуре данных, что требует их своевременного выявления и устранения. Кроме того, к системам ИИ могут адаптироваться злоумышленники, стремящиеся обмануть их или использовать в своих целях.

Поэтому член НИСКУ при необходимости обеспечивает непрерывный мониторинг эксплуатируемых ИИ-систем и реализацию процессов итеративного улучшения их качества и безопасности. Для этого необходимо:

а) Отслеживать статистику использования ИИ-сервисов и ключевые показатели качества их работы, такие как точность, полнота, актуальность, скорость реакции и т.п., в

динамике и при необходимости инициировать процедуры миграции и переобучения моделей;

b) Отслеживать обратную связь от пользователей ИИ-систем, как внутри организации клиента, так и со стороны внешних стейкхолдеров, фиксировать и анализировать их жалобы, предложения и отзывы;

d) Проводить периодический анализ принимаемых ИИ-моделями решений на предмет выявления возможных ошибок, нарушения причинно-следственных связей, дискриминации отдельных групп пользователей и т.п.;

e) Осуществлять активный поиск уязвимостей ИИ-моделей и инфраструктуры путем тестирования на проникновение, генерации состязательных примеров и т.п. как собственными силами клиента, так и с привлечением внешних экспертов по безопасности ИИ;

f) Отслеживать появление новых типов угроз и атак на ИИ-системы, уязвимостей в используемых фреймворках и библиотеках, изменения в требованиях регуляторов и своевременно разрабатывать и внедрять необходимые контрмеры;

g) Контролировать соблюдение политик и процедур, связанных с обеспечением качества, безопасности и этичности использования ИИ, проводить периодические проверки и аудиты ИИ-систем;

h) Реализовать механизмы информирования ответственных лиц и эскалации инцидентов, связанных с некорректной работой ИИ или нарушением им установленных политик, обеспечить своевременное расследование таких инцидентов и извлечение уроков.

Частота и глубина контрольных мероприятий при мониторинге ИИ-систем определяются на основе модели угроз и оценки рисков для каждого конкретного приложения ИИ. Критически важные ИИ-системы, отказ или ошибки которых могут привести к значительному ущербу, должны подвергаться постоянному мониторингу в режиме, близком к реальному времени, в то время как для менее критичных систем может быть достаточно выборочных периодических проверок.

Учитывается, что мониторинг ИИ-систем не ограничивается наблюдением за автоматически вычисляемыми метриками, а требует экспертного анализа и интерпретации накопленных данных. Модели и алгоритмы, которые кажутся стабильными и качественными с точки зрения статистических показателей, могут принимать несправедливые или неэтичные решения из-за присущих им ограничений и предубеждений.

Поэтому член НИСКУ предусматривает участие в процессе мониторинга ИИ как специалистов в соответствующих предметных областях, так и экспертов по этическим и социальным аспектам систем ИИ. По результатам мониторинга и расследования инцидентов эксплуатация отдельных моделей ИИ может быть приостановлена, функциональность ограничена, внесены коррективы в процедуры проверки качества

данных и моделей перед внедрением, доработаны регламенты взаимодействия человека и ИИ.

Такая деятельность осуществляется членом НИСКУ в случае наличия соответствующей подготовки и опыта, а также при наличии запроса от клиента.

Важной задачей консультанта является формирование культуры ответственного использования ИИ в организации клиента, вовлечение в процесс мониторинга ИИ-систем максимально широкого круга сотрудников НИСКУ.

4.8. ОБЕСПЕЧЕНИЕ ФУНКЦИОНИРОВАНИЯ ИЗОЛИРОВАННОЙ СИСТЕМЫ ИИ (ЗАКРЫТОГО КОНТУРА)

В целях защиты конфиденциальности, целостности и доступности данных клиента и третьих лиц; предотвращения утечки персональных данных, несанкционированного доступа, искажения данных; снижения рисков состязательных атак, неконтролируемого извлечения конфиденциальной информации; обеспечения управляемости, прослеживаемости и подотчётности ИИ-решений; сохранения возможности человека вмешаться в работу ИИ, приостановить её или отменить решение; член НИСКУ может использовать закрытый контур ИИ - режим функционирования ИИ-системы, при котором полный цикл обработки данных, обучения и/или дообучения, формирования выводов, поддержки или принятия решений, взаимодействия с пользователями осуществляется в изолированной среде, без неконтролируемого обмена данными, командами с внешними сетями, сервисами, неавторизованными пользователями, а также иными системами, работающих на технологиях искусственного интеллекта.

Закрытый контур может включать как архитектурную изоляцию ИИ-системы, так и контролируемый замкнутый цикл обратной связи, при котором выходы ИИ не влияют на объекты, процессы или людей без предусмотренных средств контроля, валидации и человеческого надзора.

Член НИСКУ разъясняет клиенту, что закрытый контур не отменяет и не ограничивает юридическую и этическую ответственность человека и организации за последствия применения ИИ.

Член НИСКУ обосновывает клиенту использование ИИ-системы закрытого контура, преимущества и риски ее использования, результаты тестирования и аудита, а также порядок действий при возникновении внештатных ситуаций в рамках взаимодействия сторон.

5. РАБОТА С ЗАКАЗЧИКОМ

5.1. ИНФОРМИРОВАНИЕ КЛИЕНТА О ВОЗМОЖНОСТЯХ И ОГРАНИЧЕНИЯХ ИИ

Член НИСКУ выступает в качестве надежного советника и проводника для клиента в сфере ИИ-технологий, предоставляя ему актуальную, полную и сбалансированную информацию о потенциальных выгодах, возможностях, рисках и ограничениях, связанных с использованием ИИ в его деятельности. Рекомендации консультанта основываются на детальном анализе индустрии клиента, уровня развития технологий, лучших практик, а также соображений этики и социальной ответственности. В частности, консультант осуществляет следующие действия:

a) Предоставление клиенту обзора современного состояния технологий ИИ и динамики их развития в краткосрочной, среднесрочной и долгосрочной перспективе. Обзор включает классификацию методов и инструментов ИИ, оценку их зрелости и применимости для решения различных бизнес-задач.

b) Ознакомление клиента с реальными кейсами успешного применения ИИ в его отрасли и смежных отраслях, как на глобальном, так и на локальном уровне. Кейсы обозначают предпосылки внедрения ИИ, использованные подходы, достигнутые результаты, извлеченные уроки и перспективы дальнейшего развития.

c) Проведение углубленного обучения сотрудников клиента, ответственных за принятие решений, основам ИИ, его возможностям и ограничениям с учетом специфики бизнес-процессов и данных клиента. Обучение может проводиться в форме семинаров, воркшопов, демонстраций и должно сочетать теоретические знания с практическими упражнениями. В обязательном порядке при старте каждого нового проекта должен проводиться специализированный семинар, направленный на выявление и обсуждение потенциальных этических рисков, присущих конкретному проекту.

d) Организация для руководства и сотрудников клиента посещения передовых центров компетенций, лабораторий, компаний, успешно использующих ИИ-технологии, для обмена опытом и знакомства с лучшими практиками. Такие визиты дадут возможность увидеть ИИ в действии, пообщаться с практиками и оценить потенциал ИИ для своей организации.

e) Предоставление клиенту сбалансированного анализв возможностей применения ИИ для решения конкретных задач и достижения бизнес-целей клиента. Анализ учитывает особенности доступных клиенту данных, ИТ-инфраструктуры, бизнес-процессов, корпоративной культуры и ресурсных ограничений. Обозначаются ожидаемые выгоды, необходимые инвестиции, риски и ограничения, связанные с ИИ.

f) Информирование клиента о важности качества данных для реализации проектов ИИ, провести оценку зрелости процессов управления данными клиента, дать рекомендации по организации сбора, хранения, обработки и использования данных для целей ИИ с учетом отраслевой специфики и законодательных требований.

g) Проведение детального анализа готовности организации к внедрению ИИ с точки зрения компетенций сотрудников, культуры, процессов, инфраструктуры.

h) Предоставление клиенту на регулярной основе информацию о развитии технологий и регулирования ИИ, появлении новых инструментов, платформ, открытых датасетов и моделей, изменениях в законодательстве и стандартах, связанных с использованием ИИ, обучение сотрудников навыкам работы с новыми системами ИИ.

Член НИСКУ также честно и открыто информирует клиента об ограничениях и потенциальных негативных эффектах применения ИИ-технологий, включая:

a) Вероятность ошибок и сбоев, присущую ИИ-системам в силу их вероятностной природы, и невозможность полностью исключить риски с помощью тестирования;

b) Непрозрачность и сложность интерпретации решений, принимаемых алгоритмами глубокого обучения, ограниченность практической применимости существующих подходов к объяснению таких решений;

c) Возможность дискриминации отдельных групп пользователей, усиления социального неравенства и предвзятости из-за несбалансированности обучающих выборок и исторических данных;

d) Потенциал ИИ-систем для вытеснения человеческого труда, изменений в структуре компетенций и занятости, социальных и психологических последствий для сотрудников, клиентов и общества;

e) Риски нарушения конфиденциальности, неприкосновенности частной жизни и утечек персональных данных при разработке и эксплуатации систем ИИ, необходимость соблюдения законодательных требований GDPR (General Data Protection Regulation – Общий регламент защиты данных), CCPA (California Consumer Privacy Act – Закон штата Калифорния о защите конфиденциальности) и др.;

f) Уязвимости ИИ-моделей и систем перед целевыми атаками и манипуляциями со стороны злоумышленников, которые могут перехватывать управление, деградировать качество или красть интеллектуальную собственность;

g) Негативное влияние ИИ-систем на критическую инфраструктуру, окружающую среду, физическую безопасность и здоровье людей в долгосрочной перспективе, необходимость проведения многосторонних оценок рисков и воздействий.

h) Ограниченность текущего уровня развития ИИ-технологий для решения задач, требующих здравого смысла, широких знаний, гибкости, креативности, эмпатии, моральных суждений, присущих человеческому интеллекту.

Член НИСКУ проявляет осторожность, чтобы не допускать у клиента чрезмерного оптимизма и завышенных ожиданий от ИИ, четко разделять доказанные возможности и гипотетические перспективы, ограничивать применение ИИ областями, где соотношение выгод и рисков является приемлемым с учетом специфики задач и данных клиента.

5.2. ОПРЕДЕЛЕНИЕ ЗАДАЧ И МЕТРИК ЭФФЕКТИВНОСТИ ИИ-РЕШЕНИЙ

Член НИСКУ помогает клиенту сформулировать бизнес- и технические требования к ИИ-решениям, трансформировать заявленные потребности в конкретные измеримые задачи машинного обучения, а также определить количественные показатели для оценки результатов работы ИИ-систем на всех этапах их жизненного цикла. Для этого консультант:

a) организует серии интервью и воркшопов с ключевыми стейкхолдерами клиента для сбора и приоритизации функциональных и нефункциональных требований к ИИ-системе;

b) на основе требований и анализа доступных данных (п. 3.3) формулирует конкретные задачи машинного обучения (классификация, регрессия, кластеризация и т.п.) с указанием целевых переменных, признаков и ограничений;

c) совместно с клиентом определяет набор бизнес-метрик (например, конверсия, выручка, снижение затрат) и технических метрик (точность, полнота, F1-мера и др.), увязывая их с KPI организации; при выборе метрик руководствуется принципами, изложенными в п. 3.1 (шаблон SMART);

d) согласовывает минимальные целевые значения метрик, условия успешного прохождения контрольных точек, а также порядок документирования и интерпретации результатов;

e) фиксирует все согласованные требования, задачи и метрики в техническом задании, которое служит основой для разработки, тестирования и приёмки ИИ-решения. При изменении приоритетов бизнеса или выявлении новых данных требования корректируются по согласованию с клиентом.

Член НИСКУ документирует согласованные задачи, требования и метрики в техническом задании, которое будет служить основой для последующего выбора моделей, данных и процессов разработки и оценки ИИ-решений. При изменении приоритетов бизнеса, выявлении новых данных или получении дополнительной информации от пользователей консультант оперативно корректирует исходные требования по согласованию с клиентом.

5.3. УЧЕТ РИСКОВ И ОБЯЗАТЕЛЬСТВ, СВЯЗАННЫХ С ВНЕДРЕНИЕМ ИИ

Внедрение систем ИИ в деятельность НИСКУ может создавать новые риски и приводить к возникновению обязательств, требующих тщательного анализа и выработки стратегий реагирования. Консультант помогает клиенту идентифицировать, оценить и расставить приоритеты таких рисков и обязательств, а также разрабатывает политики и процедуры по их снижению и исполнению. Для этого следует:

a) Определить область применения ИИ-решений в организации клиента и установить границы ответственности за их разработку, тестирование, эксплуатацию, обслуживание и вывод из эксплуатации. Четко распределить роли и обязанности между внутренними подразделениями клиента и внешними поставщиками услуг (включая консультанта) на каждом этапе жизненного цикла ИИ-систем.

b) Провести комплексную оценку рисков, связанных с ИИ, на основе отраслевых и внутренних данных клиента. Риски могут включать операционные сбои и потерю контроля над процессами вследствие ошибок ИИ, финансовые потери из-за неверных решений, нарушения конфиденциальности и утечки данных, дискриминацию клиентов/сотрудников, ущерб репутации, претензии регуляторов, иски о возмещении вреда и др. Для проведения оценки в обязательном порядке должна использоваться методология ISO 31010 или иная признанная международная методология. По результатам оценки должен быть составлен реестр рисков ИИ.

d) Разработать стратегии управления рисками ИИ в зависимости от их уровня и толерантности к рискам. Стратегии могут включать уклонение от риска (отказ от ИИ в определенных случаях), передачу риска (страхование, аутсорсинг), принятие риска (с установлением лимитов), снижение риска (за счет улучшения процессов разработки, тестирования, мониторинга ИИ). В рамках стратегии передачи риска консультант рекомендует и может помочь клиенту в оформлении специализированных страховых продуктов, покрывающих риски ИИ, таких как киберстрахование (cyberinsurance) и страхование профессиональной ответственности (professionalliabilityinsurance) с учетом специфики рисков ИИ.

e) Определить юридические и этические обязательства, которые накладывает на НИСКУ использование ИИ-систем, в зависимости от характера решаемых задач, типов обрабатываемых данных, отрасли и юрисдикции. Обязательства могут вытекать из законов о защите персональных данных, равенстве и недискриминации, безопасности продукции, защите прав потребителей, интеллектуальной собственности, а также добровольных этических кодексов и стандартов.

f) Разработать политики и регламенты, обеспечивающие соблюдение применимых к ИИ-системам требований на протяжении всего их жизненного цикла. Политики должны определять принципы и подходы к обеспечению прозрачности, объяснимости, недискриминационности, подотчетности ИИ-решений, стандарты аудита и верификации моделей, порядок взаимодействия человека и ИИ, механизмы апелляции решений.

g) Внедрить механизмы документирования и протоколирования процессов разработки, тестирования и использования ИИ-систем, чтобы обеспечить возможность независимой проверки их соответствия внутренним политикам и внешним требованиям. Документация должна включать спецификации данных и моделей, результаты тестирования и аудита, журналы решений и взаимодействий, инциденты и принятые меры.

h) Организовать обучение сотрудников, вовлеченных в разработку и использование ИИ, принципам этичного и ответственного ИИ, требованиям применимого законодательства, лучшим практикам индустрии. Обучение должно носить регулярный характер и обеспечивать понимание сотрудниками своих обязанностей и последствий несоблюдения установленных политик и процедур.

Учитывая быстрые темпы развития технологий и регулирования ИИ, консультант на постоянной основе отслеживает изменения в этой области и своевременно информировать клиента о появлении новых рисков, обязательств и лучших практик.

5.4. ОБЕСПЕЧЕНИЕ КОНФИДЕНЦИАЛЬНОСТИ И БЕЗОПАСНОСТИ ДАННЫХ КЛИЕНТА

При реализации проектов по внедрению ИИ консультант получает доступ к большим объемам данных клиента, в том числе конфиденциальным и персональным данным. Обеспечение защиты этих данных от несанкционированного доступа, изменения, раскрытия и уничтожения является одной из ключевых обязанностей члена НИСКУ и необходимым условием доверия в отношениях с клиентом. Для этого член НИСКУ:

a) Соблюдает применимое законодательство в области защиты данных. Внедряет в своей организации необходимые политики и механизмы контроля для обеспечения конфиденциальности, целостности и доступности данных клиентов.

b) Разрабатывает и подписывает с клиентом соглашения об уровне услуг (SLA) и обработке данных (DPA), четко определяющие ожидания и обязательства сторон в отношении безопасности данных, включая перечень обрабатываемых категорий данных, цели и сроки обработки, требования к защите данных, механизмы аудита и отчетности, штрафные санкции за нарушения и др.

c) Внедряет механизмы разграничения доступа к данным клиента на основе ролей и минимально необходимых привилегий. Доступ к конфиденциальным данным предоставляется только сотрудникам, непосредственно участвующим в проекте, после прохождения необходимого обучения и подписания соглашений о неразглашении.

e) Применяет методы анонимизации, псевдонимизации и обезличивания персональных данных там, где это возможно без ущерба для задач проекта.

g) Заключает с субподрядчиками, привлекаемыми к обработке данных клиента, соглашения, налагающие на них обязательства по обеспечению конфиденциальности и безопасности на уровне не ниже, чем у консультанта. Регулярно проверяет соблюдение субподрядчиками требований информационной безопасности. Уведомляет клиента и получает его одобрение при замене или привлечении новых субподрядчиков.

i) Уведомляет клиента обо всех запросах на раскрытие его данных от государственных органов, кроме случаев, когда это запрещено законом.

l) Организовывает обучение и повышение осведомленности сотрудников по вопросам конфиденциальности и безопасности данных. Доводит до персонала важность выполнения требований политик и процедур, ответственность за нарушения. Проводит тренировки по реагированию на инциденты информационной безопасности, в том числе с привлечением клиентов.

m) Своевременно информирует клиентов о любых фактических или подозреваемых инцидентах, связанных с безопасностью их данных, и принятых мерах по расследованию и устранению последствий.

n) По завершении проекта или по требованию клиента обеспечивает возврат и/или уничтожение всех данных в соответствии с согласованными процедурами.

Член НИСКУ позиционируется как надежный и ответственный партнер, для которого защита данных является одним из ключевых приоритетов.

5.5. ПЕРЕДАЧА ЗНАНИЙ И ОБУЧЕНИЕ КОМАНДЫ ЗАКАЗЧИКА

Успех проектов по внедрению ИИ в значительной степени зависит от способности клиента эффективно использовать и развивать внедренные решения после завершения проекта. Поэтому важной задачей члена НИСКУ, если это отражено в контракте, является передача знаний, навыков и лучших практик, полученных в ходе проекта, команде заказчика, а также содействие в формировании у клиента необходимых компетенций для самостоятельной работы с ИИ. Для этого член НИСКУ осуществляет следующие действия:

a) Включение мероприятий по передаче знаний и обучению в план проекта и выделение на них достаточных ресурсов. Согласование с клиентом ожидаемых результатов обучения, целевой аудитории, форматов, графика и бюджета учебных активностей.

b) Проведение на начальном этапе проекта оценки уровня знаний и навыков сотрудников клиента в области ИИ и науки о данных с помощью интервью, тестов, практических заданий. На основе оценки и в соответствии с целями клиента разрабатывается и согласовывается индивидуальный учебный план по управлению ИИ, охватывающий ключевые аспекты ответственного ИИ, управления данными, моделью и рисками.

c) Ознакомление команды клиента с методологией, используемой консультантом для разработки и внедрения ИИ-решений.

d) Подготовка и проведение серии технических семинаров и мастер-классов для команды клиента.

6. ПРОФЕССИОНАЛЬНОЕ РАЗВИТИЕ ЧЛЕНОВ НИСКУ

6.1. ПОВЫШЕНИЕ УРОВНЯ ЭКСПЕРТИЗЫ В ОБЛАСТИ ИИ-ТЕХНОЛОГИЙ

Члены НИСКУ, работающие с ИИ-системами, постоянно совершенствуют свои знания и навыки, чтобы предоставлять клиентам услуги самого высокого качества. Для этого консультанты осуществляют следующие действия:

a) Отслеживают последние достижения в области ИИ, машинного обучения, обработки данных, читая научные публикации, отраслевые отчеты, блоги экспертов.

b) Изучают и осваивают новые инструменты, библиотеки, фреймворки, облачные платформы для разработки и внедрения ИИ-решений.

c) Участвуют в профильных конференциях, семинарах, воркшопах, хакатонах, в том числе в качестве докладчиков и менторов.

d) Проходят онлайн-курсы, сертификационные программы, получать микро-степени от ведущих университетов и компаний отрасли.

e) Реализуют собственные проекты с использованием ИИ для закрепления полученных знаний на практике.

Члены НИСКУ, работающие с ИИ-системами могут участвовать в профильных сертификациях (IEEE – Институт инженеров электротехники и электроники, ISO – Международная организация по стандартизации).

6.2. РАЗВИТИЕ ЭТИЧЕСКИХ КОМПЕТЕНЦИЙ

Члены НИСКУ обладают не только техническими навыками, но и пониманием этических аспектов применения ИИ, умением предвидеть и оценивать потенциальные риски и социальные последствия своих рекомендаций. Для развития этических компетенций члены НИСКУ:

a) Изучают профессиональные кодексы этики для специалистов по анализу данных и ИИ, разработанные авторитетными организациями (например, ASA, ACM, IEEE).

b) Знакомятся с лучшими практиками и принципами этичного ИИ, предлагаемыми ведущими компаниями отрасли, исследовательскими институтами, некоммерческими организациями.

c) Анализируют кейсы этических дилемм и проблем, возникающих при использовании ИИ в реальных условиях.

d) Проходят специализированные тренинги по этике ИИ, организуемые НИСКУ, профессиональными ассоциациями и образовательными учреждениями.

e) Принимают участие в деятельности комитета по этике в НИСКУ и отраслевых организациях.

6.3. ОБМЕН ЗНАНИЯМИ И ЛУЧШИМИ ПРАКТИКАМИ В ПРОФЕССИОНАЛЬНОМ СООБЩЕСТВЕ

Члены НИСКУ активно участвуют в жизни профессионального сообщества, обмениваться опытом и идеями с коллегами, вносить вклад в развитие стандартов и лучших практик применения ИИ в управленческом консалтинге. Это может включать:

a) Участие в деятельности профессиональных ассоциаций консультантов, экспертных групп и рабочих органов по разработке отраслевых стандартов.

b) Выступления на конференциях и публикацию статей, описывающих результаты реализованных проектов и извлеченные уроки.

c) Менторство и наставничество для более младших коллег, студентов и начинающих специалистов.

d) Участие в сообществах практиков по внедрению ИИ в организациях, обмен кейсами и шаблонами документов.

e) Открытие исходного кода библиотек, моделей и других артефактов, разработанных в ходе проектов, для использования всем сообществом.

f) Реализация онлайн-портала для обмена лучшими практиками с открытым исходным кодом

g) Проведение регулярных мозговых штурмов для обсуждения сложных кейсов.

Инвестируя в собственное профессиональное развитие и активно взаимодействуя с коллегами, члены НИСКУ смогут предоставлять своим клиентам услуги неизменно высокого качества и быть в авангарде применения новейших достижений ИИ в управленческой практике. Это будет способствовать росту доверия бизнеса к ИИ-технологиям и консалтинговой отрасли в целом.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1 РЕКОМЕНДУЕМЫЙ ЧЕК-ЛИСТ ОЦЕНКИ СООТВЕТСТВИЯ ИИ-РЕШЕНИЯ ТРЕБОВАНИЯМ EU AI ACT

Настоящий чек-лист предназначен для верификации того, что критерии приемки (acceptancecriteria) ИИ-решения и процесс его тестирования соответствуют ключевым требованиям Регламента (ЕС) 2024/1689 (EU AI Act)[4] (далее – Регламент).

В соответствии с Регламентом все ИИ-системы делятся на категории риска. High-risk AI – это системы, перечисленные в Приложении III Регламента (например, в сферах

[4] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [Электронный ресурс]. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

критической инфраструктуры, образования, трудоустройства, доступа к государственным услугам, правоохранительной деятельности и др.). Для таких систем установлены наиболее строгие требования (статьи 8–15, 43 и др.). Системы, не относящиеся к High-risk, подлежат менее жёсткому регулированию, но обязаны соблюдать общие принципы прозрачности и информирования пользователей (статья 50). Данный чек-лист учитывает эту дифференциацию: пункты, отмеченные как «Только для High-Risk AI», применяются исключительно к системам из Приложения III Регламента; остальные пункты обязательны для всех ИИ-систем.

Чек-лист применяется до ввода системы в промышленную эксплуатацию.

Инструкция по применению:

- Ответственный: Член НИСКУ совместно с командой Заказчика.
- Основание для проверки: Документированные критерии приемки (acceptancecriteria), техническое задание, результаты тестирования, документация на систему (Model Card, Data Sheet и пр.).
- Результат: Чек-лист заполняется и подшивается к акту сдачи-приемки работ. Положительная оценка по всем пунктам, применимым к классу риска системы, является обязательным условием для принятия решения о вводе в эксплуатацию.

1. Управление рисками и соответствие

- 1.1. Разработана и внедрена система управления рисками на протяжении всего жизненного цикла ИИ-системы?
  - *Применимость: Только для High-Risk AI.*
- 1.2. Проведена классификация ИИ-системы в соответствии с приложениями I-III EU AI Act и определена ее категория риска?
  - *Применимость: Для всех систем ИИ.*

2. Качество данных и управление

- 2.1. Создана и ведется документация на набор данных (Data Sheet), описывающая его происхождение, состав, разметку и потенциальные смещения?
  - *Применимость: Только для High-Risk AI.*
- 2.2. Применяются процессы и методы для обеспечения репрезентативности, полноты и отсутствия ошибок в обучающих, валидационных и тестовых данных?
  - *Применимость: Только для High-Risk AI.*
- 2.3. Реализованы меры по устранению или смягчению предвзятости в данных, ведущей к дискриминации по защищенным признакам?
  - *Применимость: Только для High-Risk AI.*

3. Техническая документация и прозрачность

- 3.1. Подготовлена исчерпывающая техническая документация системы, включая "Model Card"?
    - *Применимость: Только для High-Risk AI.*
- 3.2. Документация содержит достаточно информации для воспроизведения системы и проведения независимого аудита?
    - *Применимость: Только для High-Risk AI.*
- 3.3. Реализована функциональность автоматической регистрации (логирования) деятельности ИИ-системы для обеспечения прослеживаемости решений?
    - *Применимость: Только для High-Risk AI.*
- 3.4. Обеспечена необходимая прозрачность для пользователей: информирование о взаимодействии с ИИ-системой, ее возможностях и ограничениях?
    - *Применимость: Для всех систем ИИ.*

4. Надзор человека и контроль

- 4.1. Предусмотрены эффективные механизмы человеческого надзора (human-in-the-loop или human-on-the-loop)?
    - *Применимость: Только для High-Risk AI.*
- 4.2. Пользователь имеет возможность игнорировать, оспорить или отменить решение, принятое системой?
    - *Применимость: Только для High-Risk AI.*

5. Точность, робастность и кибербезопасность

- 5.1. Достигнуты и задокументированы целевые показатели точности, устойчивости к ошибкам и кибербезопасности?
    - *Применимость: Только для High-Risk AI.*
- 5.2. Система протестирована на устойчивость к состязательным атакам (adversarialattacks) и концептуальному дрейфу (modeldrift)?
    - *Применимость: Только для High-Risk AI.*

6. Объяснимость (Interpretability)

- 6.1. Реализованы и задокументированы механизмы, обеспечивающие интерпретацию и объяснение решений (XAI)?
    - *Применимость: Только для High-Risk AI.*
- 6.2. Объяснения являются точными (fidelity) и понятными для целевой аудитории (пользователей, надзорных органов)?
    - *Применимость: Только для High-Risk AI.*

7. Защита прав человека и запрещенные практики

- 7.1. Проведена оценка воздействия системы на фундаментальные права человека (право на неприкосновенность частной жизни, недискриминацию и пр.)?
    - *Применимость: Для всех систем ИИ.*
- 7.2. Подтверждено, что система не использует практики, запрещенные Статьей 5 EU AI Act (напр., манипулятивное сублиминное воздействие, социальное скоринговое оценивание в широком смысле)?
    - *Применимость: Для всех систем ИИ.*

8. Декларация соответствия и маркировка

- 8.1. Для High-Risk AI систем подготовлена и подписана необходимая документация для последующего прохождения процедуры conformityassessment и нанесения маркировки CE?
    - *Применимость: Только для High-Risk AI.*

Примечания:

- "High-Risk AI": Категория систем, определенная в Приложении III EU AI Act (например, в сфере управления критической инфраструктурой, образования, занятости, доступа к публичным услугам и др.).
- Член НИСКУ обязан следить за актуальностью требований в связи с поэтапным вступлением в силу положений EU AI Act.