Кодекс этики в сфере искусственного интеллекта

Кодекс этики в сфере искусственного интеллекта (далее – Кодекс) устанавливает общие этические принципы и стандарты поведения, которыми следует руководствоваться участникам отношений в сфере искусственного интеллекта (далее – Акторы ИИ) в своей деятельности, а также механизмы реализации положений настоящего Кодекса.

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

Рекомендации настоящего Кодекса рассчитаны на системы искусственного интеллекта (далее – СИИ), применяемые исключительно в гражданских (не военных) целях.

Положения Кодекса могут быть расширены и/или конкретизированы для отдельных групп Акторов ИИ в отраслевых или локальных документах по этике в сфере ИИ с учетом развития технологий, особенностей решаемых задач, класса и назначения СИИ, уровня возможных рисков, а также специфического контекста и среды, в которой применяются СИИ.

РАЗДЕЛ I
ПРИНЦИПЫ ЭТИКИ И ПРАВИЛА ПОВЕДЕНИЯ

1. ГЛАВНЫЙ ПРИОРИТЕТ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ В ЗАЩИТЕ ИНТЕРЕСОВ И ПРАВ ЛЮДЕЙ И ОТДЕЛЬНОГО ЧЕЛОВЕКА

1.1. Человеко-ориентированный и гуманистический подход.

При развитии технологий ИИ человек, его права и свободы должны рассматриваться как наивысшая ценность. Акторы ИИ должны разрабатывать технологии, которые способствуют реализации всех потенциальных возможностей человека для достижения гармонии в социальной, экономической, духовной сфере и наивысшего расцвета личности, учитывают ключевые ценности, такие как: сохранение и развитие когнитивных способностей человека и его творческого потенциала; сохранение нравственных, духовных и культурных ценностей; содействие культурному и языковому многообразию, самобытности, сохранение традиций и устоев наций, народов, этносов и социальных групп.

Человеко-ориентированный и гуманистический подход является базовым этическим принципом и центральным критерием оценки этичного поведения Акторов в сфере ИИ, перечень которых определен в разделе 2 Кодекса.

1.2. Уважение автономии и свободы воли человека.

Акторы ИИ должны принимать необходимые меры для сохранения автономии и свободы воли человека в принятии им решений, права выбора и в целом сохранения интеллектуальных способностей человека как самостоятельной ценности и системообразующего фактора современной цивилизации. Акторы ИИ должны на этапе создания СИИ прогнозировать негативные последствия для развития когнитивных способностей человека и не

допускать разработку СИИ, которые целенаправленно вызывают такие последствия.

1.3. Соответствие закону.

Акторы ИИ должны знать и соблюдать положения законодательства Российской Федерации во всех сферах своей деятельности и на всех этапах создания, внедрения и использования технологий ИИ, в том числе в вопросах юридической ответственности Акторов.

1.4. Недискриминация.

В целях обеспечения справедливости и недопущения дискриминации Акторы ИИ должны принимать меры для того, чтобы удостовериться, что применяемые ими алгоритмы и наборы данных, методы обработки используемых для машинного обучения данных, при помощи которых осуществляется группирование и/или классификация данных, касающихся отдельных лиц или групп лиц, не влекут их дискриминацию. Акторам рекомендуется создавать и применять методики и программные решения, выявляющие и препятствующие возникновению дискриминации по признакам расовой, национальной принадлежности, политических взглядов, религиозных убеждений, возраста, половой принадлежности, социального и экономического статуса или сведений о частной жизни.

1.5. Оценка рисков и гуманитарного воздействия.

Акторам ИИ рекомендуется проводить оценку потенциальных рисков применения СИИ, включая социальные последствия для человека, общества и государства, гуманитарного воздействия СИИ на права и свободы человека на разных стадиях ее жизненного цикла, в том числе при формировании и использовании наборов данных; осуществлять долгосрочный мониторинг проявления таких рисков;

учитывать сложность поведения СИИ, включая взаимосвязь и взаимозависимость процессов в жизненном цикле СИИ при оценке рисков.

Для критических приложений СИИ 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 СИИ, а также охраны интеллектуальной собственности и коммерческой тайны разработчика.

2. НЕОБХОДИМО ОСОЗНАВАТЬ ОТВЕТСТВЕННОСТЬ ПРИ СОЗДАНИИ И ИСПОЛЬЗОВАНИИ ИИ

2.1. Риск-ориентированный подход.

Уровень внимания к этическим вопросам в области ИИ и характер соответствующих действий Акторов ИИ должен быть пропорционален оценке уровня рисков, создаваемых конкретными технологиями и СИИ для интересов человека и общества. Оценка уровня рисков учитывает как известные, так и возможные риски, при этом принимается во внимание, как уровень вероятности угроз, так и их возможный масштаб в краткосрочной и долгосрочной перспективе. Принятие значимых для общества и государства решений в области применения ИИ должно сопровождаться научно выверенным, междисциплинарным прогнозированием социально-экономических последствий и рисков, изучением возможных изменений в ценностно-культурной парадигме развития общества с учетом национальных приоритетов.

Во исполнение настоящего Кодекса рекомендуется разработка и использование единой методики оценки рисков СИИ.

2.2. Ответственное отношение.

Акторы ИИ должны ответственно относиться к вопросам влияния СИИ на общество и граждан на каждом этапе жизненного цикла СИИ, 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 СИИ, а также при выборе и использовании аппаратных средств и программного обеспечения, задействованных на различных жизненных циклах СИИ.

При этом ответственность Акторов ИИ должна соответствовать характеру, степени и размеру ущерба, который может последовать в результате использования технологий и СИИ, а также учитывать роль Актора в жизненном цикле СИИ и степень возможного и реального влияния конкретного Актора ИИ на причинение ущерба и его размер.

2.3. Предосторожность.

Когда деятельность Акторов ИИ может привести к морально неприемлемым для человека и общества последствиям, наступление которых соответствующий Актор ИИ может разумно предположить, им должны быть приняты меры, чтобы предотвратить или ограничить наступление таких последствий. Для оценки категории «моральной неприемлемости последствий» и обсуждения возможных мер их предотвращения Акторы используют положения настоящего Кодекса, в том числе механизмы, указанные в разделе 2 настоящего Кодекса.

2.4. Непричинение вреда.

Акторы ИИ не должны допускать использование технологий ИИ в целях причинения вреда, жизни и (или) здоровью человека, имуществу граждан и юридических лиц, окружающей среде. Любое использование, в том числе проектирование, разработка, тестирование, внедрение, эксплуатация СИИ, способных целенаправленно причинять

вред окружающей среде, жизни и (или) здоровью человека, имуществу граждан и юридических лиц, недопустимо.

2.5. Идентификация ИИ в общении с человеком.

Акторам ИИ рекомендуется осуществлять добросовестное информирование пользователей об их взаимодействии с СИИ, когда это затрагивает вопросы прав человека и критических сфер его жизни, и обеспечивать возможность прекратить такое взаимодействие по желанию пользователя.

2.6. Безопасность работы с данными.

Акторы ИИ должны соблюдать законодательство Российской Федерации в области персональных данных и охраняемых законом тайн при использовании СИИ; обеспечивать охрану и защиту персональных данных, обработка которых осуществляется СИИ или Акторами ИИ в целях разработки и совершенствования СИИ; разрабатывать и внедрять инновационные методы борьбы с несанкционированным доступом третьих лиц к персональным данным; использовать качественные и репрезентативные наборы данных, полученные без нарушения закона из надежных источников.

2.7. Информационная безопасность.

Акторы ИИ должны обеспечивать максимально возможную защиту от несанкционированного вмешательства в работу СИИ третьих лиц; внедрять адекватные технологии информационной безопасности, в том числе применять внутренние механизмы защиты СИИ от несанкционированных вмешательств и информирования пользователей и разработчиков о таких вмешательствах; содействовать информированию пользователей о правилах информационной безопасности при использовании СИИ.

2.8. Добровольная сертификация и соответствие положениям Кодекса.

Акторам ИИ рекомендуется внедрять системы добровольной сертификации соответствия разработанных технологий ИИ нормам, установленным законодательством Российской Федерации и настоящим Кодексом. Акторы ИИ могут создавать системы добровольной сертификации и маркировки СИИ, свидетельствующие о прохождении данными системами процедур добровольной сертификации и подтверждающих стандарты качества.

2.9. Контроль рекурсивного самосовершенствования СИИ.

Акторам ИИ рекомендуется сотрудничать в выявлении и проверке информации о способах и формах создания так называемых универсальных («сильных») СИИ и предотвращении возможных угроз, которые они несут. Вопрос применения технологий «сильного» ИИ должен находиться под контролем государства.

3. ОТВЕТСТВЕННОСТЬ ЗА ПОСЛЕДСТВИЯ ПРИМЕНЕНИЯ СИИ ВСЕГДА НЕСЕТ ЧЕЛОВЕК

3.1. Поднадзорность. Акторам ИИ следует обеспечивать комплексный надзор человека за любыми СИИ в объеме и порядке, зависящим от назначения СИИ, в том числе, например, фиксировать существенные решения человека на всех этапах жизненного цикла СИИ, или предусматривать регистрационные записи работы СИИ; обеспечивать прозрачность применения СИИ и возможность отмены человеком и (или) предотвращения принятия социально и юридически значимых решений и действий СИИ на любом этапе жизненного цикла СИИ там, где это разумно применимо.

3.2. Ответственность. Акторы ИИ не должны допускать передачи полномочий ответственного нравственного выбора СИИ, не

делегировать ответственность за последствия принятия решений СИИ – за все последствия работы СИИ всегда должен отвечать человек. Акторам ИИ рекомендуется принимать все меры для определения ответственности конкретных участников жизненного цикла СИИ с учетом их роли и специфики каждого этапа.

4. ТЕХНОЛОГИИ ИИ МОЖНО И НУЖНО ВНЕДРЯТЬ ТАМ, ГДЕ ЭТО ПРИНЕСЁТ ПОЛЬЗУ ЛЮДЯМ

4.1. Применение СИИ в соответствии с предназначением.

Акторы ИИ должны использовать СИИ в соответствии с заявленным предназначением, в предписанной предметной области, для решения предусмотренных прикладных задач.

4.2. Стимулирование развития ИИ.

Акторы ИИ должны поощрять и стимулировать разработку, внедрение и развитие безопасных и этичных решений в сфере технологий ИИ с учетом национальных приоритетов.

5. ИНТЕРЕСЫ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ ВЫШЕ ИНТЕРЕСОВ КОНКУРЕНЦИИ

5.1. Корректность сравнений СИИ.

Для поддержания добросовестной конкуренции и эффективного сотрудничества разработчиков при сравнении СИИ между собой Акторам ИИ рекомендуется использовать максимально достоверную и сравнимую информацию о возможностях СИИ применительно к задаче, а также обеспечивать единство методики измерений.

5.2. Развитие компетенций.

Акторам ИИ рекомендуется следовать принятым в профессиональном сообществе практикам, поддерживать должный уровень профессиональной компетенции, необходимый для безопасной и эффективной работы с СИИ, содействовать повышению

профессиональной компетенции работников в области ИИ, в том числе в рамках программ и образовательных дисциплин по этике ИИ

5.3. Сотрудничество разработчиков.

Акторам ИИ рекомендуется развивать сотрудничество в рамках сообщества Акторов ИИ, прежде всего разработчиков, в том числе путем информирования о выявленных критических уязвимостях с целью предотвращения их массового распространения, а также прилагать усилия для повышения качества и доступности ресурсов в сфере разработки СИИ, в том числе путем:

повышения доступности данных, в том числе размеченных;

обеспечения совместимости разрабатываемых СИИ там, где это применимо;

создания условий для формирования национальной школы развития технологий ИИ, в том числе общедоступные национальные репозитории библиотек и моделей сетей, доступные национальные средства разработки, открытые национальные фреймворки и др.;

обмена информацией о лучших практиках развития технологий ИИ;

организации или проведения конференций, хакатонов, публичных конкурсов или участия в них, школьных и студенческих олимпиад;

повышения доступности знаний и поощрения использования открытых баз знаний;

формирования условий для привлечения инвестиций в развитие технологий ИИ от российских частных инвесторов, бизнес-ангелов, венчурных фондов и фондов прямых инвестиций,

стимулирования научной, образовательной, просветительской деятельности в сфере ИИ путем участия в проектах и деятельности

ведущих научно-исследовательских центров и образовательных организаций России.

6. ВАЖНА МАКСИМАЛЬНАЯ ПРОЗРАЧНОСТЬ И ПРАВДИВОСТЬ В ИНФОРМИРОВАНИИ ОБ УРОВНЕ РАЗВИТИИ ТЕХНОЛОГИЙ ИИ, ИХ ВОЗМОЖНОСТЯХ И РИСКАХ

6.1. Достоверность информации о СИИ.

Акторам ИИ рекомендуется предоставлять пользователям СИИ достоверную информацию о СИИ, допустимых областях и наиболее эффективных методах применения СИИ, вреде, пользе и существующих ограничениях в их применении.

6.2. Повышение осведомлённости об этике применения ИИ.

Акторам ИИ рекомендуется проводить мероприятия, направленные на повышение уровня доверия и осведомлённости граждан, являющихся пользователями СИИ в частности, и общества в целом, о разрабатываемых технологиях, особенностях этичного применения СИИ и иных сопутствующих развитию СИИ положениях всеми доступными способами, в том числе путём разработки научных, публицистических материалов, организации научных и общественных конференций, семинаров, а также посредством включения в правила эксплуатации СИИ правил этичного поведения пользователей и (или) эксплуатантов.

РАЗДЕЛ 2
ПРИМЕНЕНИЕ КОДЕКСА

1. ОСНОВЫ ДЕЙСТВИЯ КОДЕКСА

1.1. Правовая основа Кодекса.

Кодекс учитывает законодательство Российской Федерации, в том числе Конституцию Российской Федерации, иные нормативно-правовые акты, документы стратегического планирования, включая Национальную стратегию развития искусственного интеллекта, Стратегию национальной безопасности Российской Федерации, Концепцию регулирования искусственного интеллекта и робототехники, а также ратифицированные Российской Федерацией международные договоры и соглашения, применимые к вопросам обеспечения прав и свобод граждан в контексте использования информационных технологий.

1.2. Термины.

Термины и определения в настоящим Кодексе определяются в соответствии с действующими нормативными правовыми актами, документами стратегического планирования и нормативно-технического регулирования в сфере ИИ.

1.3. Акторы ИИ.

Для целей настоящего Кодекса под Акторами ИИ понимается круг лиц, в том числе иностранных, принимающих участие в жизненном цикле СИИ при его реализации на территории Российской Федерации или в отношении лиц, находящихся на территории Российской Федерации, включая предоставление товаров и оказание услуг. К таким лицам относятся, в том числе:

разработчики, создающие, обучающие, тестирующие модели/системы ИИ и разрабатывающие, реализующие такие модели/системы программные и/или аппаратные комплексы и принимающие на себя ответственность в отношении их конструкции;

заказчики (организация или лицо), получающие продукт или услугу;

поставщики данных и лица, осуществляющие формирование наборов данных для применения их в СИИ;

эксперты, осуществляющие измерение и/или оценку параметров разработанных моделей/систем;

изготовители, осуществляющие производство СИИ;

эксплуатанты СИИ, на законном основании владеющие соответствующими системами, использующие их по назначению и непосредственно реализующие решение прикладных задач с использованием СИИ;

операторы (лицо или организация), осуществляющие работу СИИ;

лица, принимающие участие в регуляторном воздействии на сферу ИИ, в том числе разработчики нормативно-технических документов, руководств, различных регуляторных положений, требований и стандартов в области ИИ;

иные лица, действия которых потенциально могут повлиять на результаты действий СИИ или лиц, принимающих решения с использованием СИИ.

2. МЕХАНИЗМ ПРИСОЕДИНЕНИЯ И РЕАЛИЗАЦИИ КОДЕКСА

2.1.Добровольность присоединения.

Присоединение к Кодексу является добровольным. Присоединяясь к Кодексу, Акторы ИИ на добровольной основе соглашаются следовать его рекомендациям.

Присоединение и следование положениям настоящего Кодекса может приниматься во внимание при предоставлении мер поддержки или ином взаимодействии с Актором ИИ или между Акторами ИИ.

2.2.Уполномоченные по этике и/или комиссии по этике.

Для обеспечения реализации положений настоящего Кодекса и действующих правовых норм при создании, применении и использовании СИИ Акторы ИИ назначают уполномоченных по этике ИИ, ответственных за реализацию Кодекса и являющихся контактными лицами Актора ИИ по вопросам этики ИИ, а также могут создавать коллегиальные отраслевые органы – внутренние комиссии по этике в сфере ИИ для рассмотрения наиболее актуальных или спорных вопросов в сфере этики ИИ. Акторам ИИ рекомендуется определять уполномоченного по этике ИИ по возможности при присоединении к настоящему Кодексу или в течение двух месяцев с момента присоединения к Кодексу.

2.3. Комиссия по реализации Национального кодекса в сфере этики ИИ.

В целях практической и всесторонней имплементации принципов этики и правил поведения в сфере ИИ создается Комиссия по реализации Кодекса в сфере этики ИИ (далее – Комиссия), обеспечивающая участие в ее работе Акторов ИИ, присоединившихся к настоящему Кодексу.

Комиссия может иметь рабочие органы и группы, состоящие из представителей бизнес-сообщества, науки, государственных органов и иных заинтересованных сторон. В рамках Комиссии рассматриваются заявления Акторов ИИ на присоединение к положениям настоящего Кодекса, наиболее актуальные или спорные вопросы в сфере этики ИИ, утверждается свод наилучших практик, осуществляется разработка и утверждение методик, а также обеспечивается реализация других положений настоящего Кодекса.

Обеспечение деятельности Комиссии и ведение ее секретариата осуществляется Ассоциацией «Альянс в сфере искусственного интеллекта» при участии иных заинтересованных организаций.

2.4. Реестр участников Кодекса.

Для присоединения к настоящему Кодексу Актор ИИ направляет упомянутую Комиссию соответствующее заявление. Реестр Акторов ИИ, присоединившихся к Кодексу, является публичным и ведется на общедоступном сайте/портале.

2.5. Разработка методик и руководств.

Для реализации Кодекса рекомендуется разработка методик, руководств, «чек-листов» и иных методологических материалов, обеспечивающих наиболее эффективное соблюдение положений Кодекса Акторами ИИ.

2.6. Свод практик.

В целях своевременного обмена передовым опытом полезного и безопасного применения СИИ, построенного на базовых принципах настоящего Кодекса, повышения прозрачности деятельности разработчиков и поддержания здоровой конкуренции на рынке СИИ, Акторы ИИ могут создавать свод наилучших и/или наихудших практик решения возникающих этических вопросов в жизненном цикле ИИ, отбираемых по критериям, установленным профессиональным сообществом, и обеспечивая публичный доступ к данному своду практик.