

# **Национальный кодекс этики в сфере искусственного интеллекта**

Настоящий национальный кодекс этики в сфере этики искусственного интеллекта (далее – Кодекс) устанавливает общие этические принципы и стандарты поведения, которыми могут руководствоваться участники 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.

## **Раздел 1. Принципы этики и правила поведения**

### **Главный приоритет развития технологий ИИ в защите интересов людей, отдельных групп, каждого человека.**

**1.1. Человеко-ориентированный и гуманистический подход.** При развитии технологий ИИ человек рассматривается как наивысшая ценность. Технологии ИИ должны способствовать реализации всех потенциальных возможностей человека для достижения гармонии в социальной, экономической, духовной сфере и наивысшего расцвета личности, учитывать 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, сохранению традиций и устоев наций, народов, этносов и социальных групп; поддержание взаимовыгодного, равноправного сотрудничества на региональном и международном уровнях в сфере трансфера технологий и обмена данными; содействие достижению Целей в области устойчивого развития ООН.

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

**1.2. Проектируемое соответствие закону** Акторы ИИ должны знать, учитывать и ориентироваться на положения законодательства РФ во всех сферах своей деятельности, в том числе вопросах юридической ответственности Акторов.

**1.3. Недискриминация.** В целях обеспечения справедливости и недопущения дискриминации Акторы ИИ должны принимать доступные меры для того, чтобы

удостовериться, что применяемые ими 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 умышленную дискриминацию; Акторам рекомендуется создавать и применять методики и программные решения, выявляющие и препятствующие возникновению дискриминации по признакам расовой, национальной принадлежности, политических взглядов, религиозных или философских убеждений или сведений о интимной жизни.

**1.4. Оценка рисков и гуманитарного воздействия.** Рекомендуется проводить оценку потенциальных рисков применения СИИ и 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 определяться в соответствии с оценкой 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. Все положения настоящего Кодекса должны применяться с учетом риск-ориентированного подхода, учитывая пропорциональность действий Акторов ИИ масштабу и значимости применения ими технологий ИИ. Для исполнения 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, этическое, 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 систем ИИ, 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. Принцип предосторожности.** Когда деятельность Акторов ИИ может привести к морально неприемлемым последствиям, наступление которых соответствующий Актор ИИ может разумно предположить в конкретных обстоятельствах, им должны быть приняты меры, чтобы предотвратить или ограничить наступление таких последствий. Для оценки категории «моральной приемлемости последствий» и обсуждения возможных мер его предотвращения Актеры используют положения настоящего Кодекса, в том числе механизмы, указанные в разделе IV настоящего Кодекса.

**2.4. Непричинение вреда.** Актеры ИИ не должны допускать использование технологий искусственного интеллекта 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 и уважение автономии воли человека.** Акторам ИИ там, где это разумно применимо, рекомендуется осуществлять добросовестное информирование пользователей об их взаимодействии с системой ИИ, и обеспечивать возможность прекратить такое взаимодействие по желанию пользователя.

**2.6. Безопасность работы с данными.** Актеры ИИ должны гарантировать соблюдение законодательства 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 персональными данным; использовать качественные и репрезентативные наборы данных, полученные без нарушения закона из надежных источников;

**2.7. Информационная безопасность.** Актеры ИИ должны обеспечивать максимально возможную защиту от несанкционированного вмешательства в работу СИИ третьих лиц; внедрять наиболее целесообразные технологии информационной

безопасности, в том числе там, где это разумно применимо, имплементировать 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.**

Поощряется внедрение систем добровольной сертификации соответствия разработанных технологий искусственного интеллекта нормам, установленным законодательством Российской Федерации и настоящим Кодексом. Отраслевые объединения или группы разработчиков 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.** Акторы ИИ будут сотрудничать в выявлении и проверке информации о способах и формах создания так называемых универсальных («сильных») систем ИИ и предотвращении возможных угроз, которые они несут.

## **3. Ответственность за последствия применения ИИ всегда лежит на человеке**

**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 юридически значимых решений и действий СИИ на любом этапе жизненного цикла СИИ там, где это разумно применимо и не приведет к нарушению прав и законных интересов граждан.

**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 отвечает 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 будут поощрять и стимулировать разработку, внедрение и развитие безопасных и этических решений в сфере технологий ИИ для достижений целей устойчивого развития;

## **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 необходимо 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.

**5.3. Сотрудничество разработчиков.** Акторы ИИ будут 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;
- обмена информацией о лучших практиках развития технологий искусственного интеллекта;
- 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 искусственного интеллекта от российских частных инвесторов, бизнес-ангелов, венчурных фондов и фондов прямых инвестиций,

— стимулирования научной, образовательной, просветительской деятельности в сфере искусственного интеллекта 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 пользователю достоверную информацию о допустимых областях и наиболее эффективных методах применения СИИ и существующих ограничениях.

**6.2. Повышение осведомлённости об этике применения ИИ.** Акторам ИИ рекомендуется проводить научные, студенческие, общественные и иные мероприятия, направленные на повышение уровня доверия и осведомлённости граждан, являющихся пользователями СИИ в частности и общества в целом, о разрабатываемых технологиях, вреде и пользе применения СИИ, 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.

## **7. Принципы этики развиваются по мере появления новых знаний, новых вызовов, новых возможностей.**

**7.1. Развитие принципов.** Положения настоящего Кодекса должны пересматриваться с учетом быстрого развития технологий ИИ и развития представления об этике использования технологий ИИ в обществе.

## **Раздел 2. Применение Кодекса.**

### **1. Сфера действия Кодекса**

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

**1.2. Термины.** Термины и определения в настоящем Кодексе определяются в соответствии с действующими нормативными правовыми актами и документами стратегического планирования в сфере ИИ, а также документами нормативно-технического регулирования в сфере ИИ.

**1.3. Акторы в сфере ИИ.** Для целей настоящего Кодекса под Акторами ИИ в сфере 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.

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

**1.4. Соотношение Кодекса с нормативными правовыми актами.** В основе Кодекса лежит законодательство Российской Федерации, в том числе Конституция Российской Федерации, иные нормативно-правовые акты, а также ратифицированные Российской Федерацией международные договоры и соглашения, применимые к

вопросам обеспечения прав и свобод граждан в контексте использования информационных технологий, в том числе цели ООН в области устойчивого развития; Всеобщая декларация прав человека, , Конвенция о защите прав человека и основных свобод.

## **2. Механизм реализации Кодекса**

**2.1. Уполномоченные по этике и комиссии по этике.** Для обеспечения реализации положений настоящего Кодекса и обеспечения соблюдения положений настоящего Кодекса и действующих правовых норм при создании, применении и использовании СИИ Акторы ИИ назначают уполномоченных по этике искусственного интеллекта, ответственных за реализацию Кодекса и являющихся контактными лицами Актора ИИ по вопросам этики ИИ, а также могут создавать коллегиальные отраслевые органы - 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 при присоединении к настоящему Кодексу или в течение двух месяцев с момента присоединения к Кодексу.

**2.2. 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 его публичный доступ. Присоединение и следование положениям настоящего Кодекса может также приниматься во внимание при предоставлении мер поддержки или ином взаимодействии с Актором ИИ или между Акторами ИИ.

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

**2.4. Национальная комиссия по этике ИИ.** Для обеспечения реализации положений настоящего Кодекса создается Национальная комиссия по этике ИИ (далее – Комиссия по этике ИИ), обеспечивающая участие в ее работе Акторов ИИ, присоединившихся к настоящему Кодексу. Комиссия по этике ИИ может иметь рабочие органы и группы. Обеспечение деятельности Комиссии по этике ИИ и ведение ее секретариата осуществляется Ассоциацией «Альянс в сфере



искусственного интеллекта». В рамках Комиссии по этике ИИ 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.

**2.5. Реестр присоединений к Кодексу.** Для присоединения к настоящему Кодексу Актор ИИ направляет Комиссию по этике ИИ соответствующее заявление. Реестр Акторов ИИ, присоединившихся к Кодексу, является публичным и ведется на сайте Комиссии по этике ИИ.