**Национальный кодекс этики в сфере искусственного интеллекта**

**I. Общие положения**

**1. Назначение кодекса и цели.**

Принимая во внимание динамичное развитие технологий искусственного интеллекта (далее – ИИ), а также общественную значимость результатов применения технологий ИИ,

признавая важность ответственного отношения к разработкам в области искусственного интеллекта,

учитывая потенциально высокую автономность систем ИИ (способность самостоятельного принятия решений), в частности отсутствие в ряде случаев способов точного выявления и описания их алгоритмов функционирования и подтверждения заявленных свойств,

исходя из необходимости формирования этических ориентиров поведения в связи с возникновением новых вызовов и угроз, связанных с ситуациями морального выбора/этическими дилеммами при создании, применении, использовании и регулировании технологий ИИ,в целях обеспечения доверия пользователей, общества и государства к технологиям искусственного интеллекта, а также разработки и использования технологий ИИ в интересах человека и общества на основе сбалансированного подхода и социальной ответственности бизнеса, а также повышения доверия общества к системам ИИ (далее – СИИ) и результатам их применения,

настоящий Кодекс этики в сфере искусственного интеллекта (далее – Кодекс) устанавливает общие принципы и стандарты поведения, которыми могут руководствоваться акторы в сфере искусственного интеллекта (далее – Акторы ИИ) в своей деятельности, а также механизмы реализации положений настоящего Кодекса. Действие настоящего Кодекса распространяется исключительно на СИИ, применяемые в гражданских (не военных) целях.

**2. Допустимость отраслевых кодексов.** Положения Кодекса могут быть расширены и конкретизированы для отдельных групп Акторов ИИ с учетом особенностей решаемых задач, класса и назначения СИИ, уровня возможных рисков, а также специфического контекста и среды, в которой применяются СИИ.

**3. Рекомендательный характер.** Положения настоящего Кодекса носят рекомендательный характер и применяются соответствующими Акторами ИИ на добровольной основе Этические нормы могут играть важную роль в тех случаях, где нормативные требования не сформулированы и законодательно не закреплены.

**4. Кодекс и нормативные акты.**

Кодекс направлен на достижение целей, предусмотренных Национальной стратегией развития искусственного интеллекта до 2030 г., утвержденной Указом Президента Российской Федерации
от 10 октября 2019 г. № 490, Концепцией развития регулирования отношений
в сфере технологий искусственного интеллекта и робототехники до 2024 года, утвержденной Распоряжением Правительства Российской Федерации
от 19 августа 2020 г. № 2129-р., Стратегией развития информационного общества в Российской Федерации на 2017 – 2030 годы, федеральным проектом «Искусственный интеллект».

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[[1]](#footnote-1); Всеобщая декларация прав человека[[2]](#footnote-2), Европейская конвенция по правам человека, Конвенция о защите прав человека и основных свобод [[3]](#footnote-3).

**5. Круг действия.** Для целей настоящего Кодекса под Акторами в сфере искусственного интеллекта 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. Каждый из Акторов ИИ, в том или ином виде должен понимать и принимать во внимание этическую сторону своих решений или действий с использованием СИИ или в отношении СИИ. К таким лицам (Акторы ИИ) относятся, в том числе:

- 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;

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

- 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;

- 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 (роль оператора и роль пользователя могут возлагаться одновременно или последовательно на одно и то же лицо или организацию);

- пользователи (лицо или группа лиц, в том числе потребители), извлекающие пользу из СИИ в процессе ее применения и являющиеся потребителями услуг, сервисов и продуктов, использующих технологии ИИ;

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

**6. Термины**

Термины и определения в настоящим Кодексе определяются в соответствие с действующими нормативными-правовыми актами и документами нормативно-технического регулирования в сфере ИИ[[4]](#footnote-4).

**II. Общие этические принципы / ценности поведения Акторов ИИ**

Акторы ИИ в своей деятельности руководствуются следующими общими принципами поведения:

**7.** **Человеко-ориентированный и гуманистический подход.** При развитии технологий ИИ человек рассматривается как наивысшая ценность. Технологии и СИИ 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.

Соблюдение и продвижение интересов человека является центральным критерием оценки этичного поведения Акторов в сфере ИИ и базовым принципом поведения Акторов ИИ

**8.** **Риск-ориентированный подход**. Уровень внимания к этическим вопросам в области ИИ должен определяться в соответствии с оценкой уровня рисков, создаваемых конкретными технологиями и СИИ для интересов человека и общества. Оценка уровня рисков в условиях частичной или значительной неопределенности 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.

**9.** **Принцип предосторожности.** Когда деятельность Акторов ИИ может привести к морально неприемлемому вреду, наступление которого соответствующий Актор ИИ может разумно предположить в конкретных обстоятельствах, им должны быть приняты меры, чтобы предотвратить или ограничить этот вред. Для оценки степени вреда и обсуждения возможных мер его предотвращения Акторы могут пользоваться механизмами, указанными в разделе IV настоящего Кодекса.

**10.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.

**III. Этические стандарты поведения акторов в сфере ИИ**

 Акторам ИИ необходимо соблюдать общепринятые этические нормы в области практического использования СИИ, действуют, обеспечивая соблюдение законодательства и международных обязательств Российской Федерации на всех этапах жизненного цикла СИИ, и придерживаются следующих стандартов поведения:

***A. Действовать ответственно и добросовестно***

**11. Недискриминация.** В целях обеспечения справедливости и недопущения дискриминации Акторы ИИ должны принять 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 влечет их умышленную дискриминацию по признакам расовой, национальной принадлежности, политических взглядов, религиозных или философских убеждений или сведений об интимной жизни;

**12. Непричинение вреда.** Не допускать использование технологий искусственного интеллекта в целях причинения вреда окружающей среде, жизни
и(или) здоровью человека, имуществу граждан и юридических лиц. Любое использование, в том числе проектирование, разработка, внедрение, эксплуатация систем искусственного интеллекта, способных
независимо от человека целенаправленно причинять вред окружающей среде, жизни и(или) здоровью человека, имуществу граждан и юридических лиц, недопустимо.

**13.** **Безопасность работы с данными.** Гарантировать соблюдение законодательства Российской Федерации в области персональных данных при использовании СИИ; обеспечивать охрану и защиту персональных данных, обработка которых осуществляется СИИ или Акторами в целях разработки и совершенствования СИИ; разрабатывать и внедрять инновационные методы борьбы с несанкционированным доступом третьих лиц к персональными данным; использовать наборы данных, полученные без нарушения закона, для целей развития технологий ИИ.

**14. Поднадзорность.** Обеспечивать комплексный надзор человека за любыми СИИ в объеме и порядке, зависящих от назначения СИИ, в том числе, например, фиксировать существенные решения человека на всех этапах жизненного цикла СИИ, или предусматривать регистрационные записи работы СИИ, или определять ответственность конкретных участников за каждый этап разработки и эксплуатации СИИ.

**15. Подконтрольность.** Обеспечивать 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.

**16.** **Ответственность.** Не допускать передачи полномочий ответственного нравственного выбора системам ИИ; не делегировать ответственность за последствия принятия решений СИИ – за все последствия работы СИИ всегда отвечает человек.

***Б. Соблюдать интересы общества и граждан***

**17.** **Уважение прав человека**. Соблюдать права и свободы человека, его честь и достоинство, уважать автономию воли человека;

**18.. Идентификация ИИ в общении с человеком**. При использовании технологий искусственного интеллекта пользователями там, где это разумно применимо, осуществлять добросовестное информирование пользователей о факте предстоящего использования продуктов и услуг на основе технологий искусственного интеллекта.

***В. Проявлять профессионализм и компетентность***

**19. Развитие компетенций.** Следовать принятым в профессиональном сообществе практикам, поддерживать должный уровень профессиональной компетенции, необходимый для безопасной и эффективной работы с системами искусственного интеллекта; содействовать повышению профессиональной компетенции своих работников в области искусственного интеллекта

**20.**  **Применение СИИ в соответствие с предназначением .** Использовать СИИ в соответствии с заявленным предназначением, в предписанной предметной области, для решения предусмотренных прикладных задач;

**21.  Стимулирование развития ИИ.** Поощрять и стимулировать разработку, внедрение и развитие решений в сфере технологий ИИ для достижений целей устойчивого развития;

**22. Сотрудничество разработчиков**. 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.

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;

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

Стимулировать научную, образовательную, просветительскую деятельность в сфере искусственного интеллекта путем участия в проектах и деятельности ведущих научно-исследовательских центров и образовательных организаций России.

**23. 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 ИИ и порядка работы с персональными данными и информацией;

**24. Корректность сравнений СИИ.** Для сравнения СИИ между собой использовать максимально достоверную в конкретных обстоятельствах и сравнимую информацию о возможностях СИИ применительно к задаче, особенно в условиях конкуренции решений;

***Г. Содействовать доверию общества к технологиям ИИ в интересах устойчивого развития***

**25. Достоверность информации о СИИ.** Предоставлять пользователю достоверную информацию о допустимых областях применения СИИ и существующих ограничениях.

**26.** **Доведение правил этики до пользователей** Акторам, осуществляющим создание, внедрение, введение в оборот систем искусственного интеллекта, рекомендуется предусматривать в условиях эксплуатации СИИ правила этического поведения пользователей и/или эксплуатантов, соответствующие положениям Кодекса.

**27. Информирование общества об этике применения ИИ** Способствовать информированности и осведомленности общества и пользователей о технологиях ИИ и результатах их применения, а также об этических проблемах их использования.

**IV. Механизмы реализации Кодекса**

**28.** **Оценка рисков**. Рекомендуется проводить оценку потенциальных рисков применения СИИ на разных стадиях ее жизненного цикла, учитывать сложность поведения СИИ, включая взаимосвязь и взаимозависимость процессов в жизненном цикле СИИ при оценке рисков.Для критических приложений СИИ в особых случаях приветствуются 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.

**29.** **Уполномоченный по этике и комиссии.** Для обеспечения реализации положений настоящего Кодекса и обеспечения соблюдения действующих правовых норм при создании, применении и использовании СИИ Акторы могут также использовать институт уполномоченных по этике искусственного интеллекта и/или формировать коллегиальные внутренние или внешние отраслевые органы (комиссии по этике использования технологий искусственного интеллекта).

**30**. **Свод наилучших практик и реестр.** В целях своевременного обмена передовым опытом полезного и безопасного применения СИИ, построенного на базовых принципах настоящего Кодекса, Акторы ИИ могут создавать свод наилучших практик решения возникающих этических вопросов в жизненном цикле ИИ, обеспечивая его публичный доступ. Также дополнительно в целях повышения прозрачности деятельности разработчиков и поддержания здоровой конкуренции на рынке СИИ, Акторы ИИ вправе принять решение о создании реестра добросовестных разработчиков, отбираемых по критериям, установленным профессиональным предпринимательским и научным сообществом.

31. **Повышение осведомлённости об этике применения 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.

32. **Добровольная сертификация.** 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 систем искусственного интеллекта (знак качества), свидетельствующие о прохождении данными системами процедур добровольной сертификации и подтверждающих стандарты качества.

33. **Пересмотр положений Кодекса.** Положения настоящего кодекса должны пересматриваться с учетом быстрого развития технологий ИИ и развития представления об этике использования технологий ИИ в обществе.

Приложение №1

Термины и определения, используемые в Кодексе этики в сфере искусственного интеллекта, применяются в понимании, закреплённом в следующих документах нормативно-технического регулирования в сфере ИИ:

* ГОСТ Р ИСО 26000-2012, «Руководство по социальной ответственности», идентичен международному стандарту ISO 26000:2010 "Guidance on social responsibility" (раскрывает основные проблемы и принципы социальной ответственности, содержит определения терминов «подотчетность», «должная предусмотрительность», «этичное поведение», «прозрачность» и др.);
* ГОСТ Р 54089-2010. «Интегрированная логистическая поддержка. Электронное дело изделия. Основные положения» (содержит определения «разработчики», «изготовители», «эксплуатанты»);
* ГОСТ Р ИСО 9001-2015 Системы менеджмента качества. Требования. Переиздание, февраль 2020 г., идентичен международному стандарту ISO 9001:2015 "Quality management systems - Requirements", IDT (приведены принципы менеджмента качества, раскрыта концепция «риск-ориентированного мышления»);
* ГОСТ Р ИСО 31000—2019 «Менеджмент риска: принципы и руководство», идентичен международному стандарту ISO 31000:2018 "Risk management - Guidelines", IDT (содержит определения «риск», «менеджмент риска», «управление риском»);
* ГОСТ Р ИСО 9000-2015 Системы менеджмента качества. Основные положения и словарь (Издание с Поправкой), 6 издание, октябрь 2019 г. (раскрыты понятия «Качество», «Система менеджмента качества», «Принятие решений, основанное на свидетельствах»);
* ГОСТ Р ИСО/МЭК 12207-2010 Информационная технология. Системная и программная инженерия. Процессы жизненного цикла программных средств (определения «жизненный цикл», «аудит», «заказчик», «разработчик», «оценивание», «исполнитель», «система»);
* ГОСТ Р 53625-2009 (ИСО/МЭК 19796-1:2005) Информационная технология (ИТ). Обучение, образование и подготовка. Менеджмент качества, обеспечение качества и метрики. Часть 1. Общий подход. Переиздание, октябрь 2018 г. (раскрыт общий подход к качеству на основе основополагающих принципов и требований);
* ГОСТ Р 53723-2009 Информационные технологии (ИТ). Руководство по применению ГОСТ Р 53625-2009, Переиздание, октябрь 2018 г. (раскрывает подходы к обеспечению качества электронного обучения, приведены «эталонные критерии качества (ЭКК)»);
* ГОСТ Р ИСО/МЭК ТО 19791-2008. Информационная технология. Методы и средства обеспечения безопасности. Оценка безопасности автоматизированных систем (определения «менеджмент рисков», «верификация»);
* ГОСТ Р 52653-2006 Информационно-коммуникационные технологии в образовании. Термины и определения. Переиздание, октябрь 2018 г. (приведены определения терминов «информация», «информационная технология», «информационная система»);
* ГОСТ Р ИСО/МЭК 9126-93 Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению (определение «программно-аппаратные средства», «качество», «уровень качества функционирования»);
* ГОСТ 20886-85 Группа Э00 Организация данных в системах обработки данных. Термины и определения (определения: «набор данных», «защита данных»)
1. https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/ [↑](#footnote-ref-1)
2. https://www.un.org/ru/documents/decl\_conv/declarations/declhr.shtml [↑](#footnote-ref-2)
3. https://www.coe.int/ru/web/compass/the-european-convention-on-human-rights-and-its-protocols [↑](#footnote-ref-3)
4. См. приложение №1 [↑](#footnote-ref-4)