Банк России

КОДЕКС ЭТИКИ В СФЕРЕ РАЗРАБОТКИ И ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ФИНАНСОВОМ РЫНКЕ

1. ЦЕЛИ И ПРИНЦИПЫ РАЗРАБОТКИ И ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ФИНАНСОВОМ РЫНКЕ

1.1. Целями настоящего Кодекса являются:

1) повышение доверия физических и юридических лиц (далее – клиенты) к применению искусственного интеллекта кредитными организациями, иностранными банками, осуществляющими деятельность на территории Российской Федерации через свои филиалы, некредитными финансовыми организациями, лицами, оказывающими профессиональные услуги на финансовом рынке, субъектами национальной платежной системы (далее – организации) при оказании клиентам услуг;

2) содействие развитию искусственного интеллекта на финансовом рынке и применению организациями доверенных технологий искусственного интеллекта;

3) минимизация рисков, связанных с разработкой и применением искусственного интеллекта на финансовом рынке (далее – риски искусственного интеллекта).

1.2. Для осуществления указанных целей при разработке и применении искусственного интеллекта организациям рекомендуется соблюдать следующие принципы:

1) человекоцентричность;

2) справедливость;

3) прозрачность;

4) безопасность, надежность и эффективность;

5) ответственное управление рисками.

1.3. Организациям рекомендуется обмениваться опытом, а также участвовать в разработке совместных документов и аналитических материалов, содействующих достижению целей настоящего Кодекса.

1.4. В случае привлечения при разработке и применении искусственного интеллекта организациями лиц, в отношении которых Банк России не осуществляет контроль (надзор), организациям рекомендуется обеспечивать соблюдение такими лицами настоящего Кодекса.

2. ПРИНЦИП ЧЕЛОВЕКОЦЕНТРИЧНОСТИ

2.1. Организациям рекомендуется принимать следующие меры:

1) повышение качества оказания услуг;

2) предоставление клиентам возможности отказаться от взаимодействия с применением искусственного интеллекта;

3) обеспечение возможности пересмотра решения, принятого с применением искусственного интеллекта;

4) повышение осведомленности клиентов о принятии решений с применением искусственного интеллекта;

5) повышение финансовой доступности.

2.2. В рамках повышения качества оказания услуг организациям рекомендуется осуществлять оценку удовлетворенности клиентов, которым оказаны услуги с применением искусственного интеллекта, а также осуществлять контроль качества оказания таких услуг.

2.3. В рамках предоставления клиентам возможности отказаться от взаимодействия с применением искусственного интеллекта организациям рекомендуется предоставить клиентам возможность взаимодействовать с сотрудником организации.

2.4. В рамках обеспечения возможности пересмотра решения, принятого с применением искусственного интеллекта, организациям рекомендуется на основании запроса клиента организовать пересмотр такого решения сотрудником организации.

2.5. В рамках повышения осведомленности клиентов о принятии решений с применением искусственного интеллекта организациям рекомендуется разъяснять действия, которые клиенту необходимо совершить для принятия искусственным интеллектом решения, на основании которого делается вывод о возможности оказания услуг. Организация вправе не предоставлять указанные разъяснения, если имеются разумные основания полагать, что такое информирование может снизить эффективность применения искусственного интеллекта.

2.6. В рамках повышения финансовой доступности организациям рекомендуется при разработке и применении искусственного интеллекта учитывать факторы уязвимости клиентов (возраст, образование, ограниченные возможности и другие) и возможное влияние указанных факторов на оказание таким клиентам услуг.

3. ПРИНЦИП СПРАВЕДЛИВОСТИ

3.1. Организациям рекомендуется принимать следующие меры:

1) обоснованность использования персональных данных о клиентах;

2) недопустимость использования дискриминационных факторов;

3) обеспечение недискриминационного характера набора данных.

3.2. В рамках обоснованности использования персональных данных о клиентах организациям рекомендуется при разработке искусственного интеллекта использовать персональные данные клиентов только в случаях, когда их использование требуется для существенного повышения эффективности применения искусственного интеллекта при оказании услуг.

3.3. В рамках недопустимости использования дискриминационных факторов организациям рекомендуется при принятии решений с применением искусственного интеллекта не учитывать, в частности, следующие факторы:

1) национальная, языковая и расовая принадлежность;

2) принадлежность к политическим партиям, общественным объединениям;

3) вероисповедание и отношение к религии.

3.4. В рамках обеспечения недискриминационного характера наборов данных организациям рекомендуется при разработке искусственного интеллекта проверять полноту, актуальность и достоверность данных о социальных группах клиентов, при оказании услуг которым организация вправе применять искусственный интеллект.

4. ПРИНЦИП ПРОЗРАЧНОСТИ

4.1. Организациям рекомендуется принимать следующие меры:

1) информирование о применении искусственного интеллекта;

2) раскрытие информации об искусственном интеллекте;

3) маркировка информации, созданной с применением больших генеративных моделей.

4.2. В рамках информирования о применении искусственного интеллекта организациям рекомендуется сообщать клиентам о применении искусственного интеллекта при оказании им услуг, если применение искусственного интеллекта неочевидно из обстоятельств.

4.3. В рамках раскрытия информации об искусственном интеллекте организациям рекомендуется предоставлять клиентам полную и достоверную информацию о рисках искусственного интеллекта и условиях применения организацией искусственного интеллекта.

4.4. В рамках маркировки информации, созданной с применением больших генеративных моделей, организациям рекомендуется маркировать сведения, в том числе изображения, аудио- и видеоматериалы, созданные с применением больших генеративных моделей, за исключением следующих случаев:

1) большие генеративные модели применяются только для редактирования сведений, которое не влечет существенного изменения содержания указанных сведений;

2) применение больших генеративных моделей очевидно из обстоятельств;

3) отсутствует риск причинения вреда клиентам.

5. ПРИНЦИП БЕЗОПАСНОСТИ, НАДЕЖНОСТИ И ЭФФЕКТИВНОСТИ

5.1. Организациям рекомендуется принимать следующие меры:

1) проверка качества наборов данных;

2) проверка качества искусственного интеллекта;

3) мониторинг качества искусственного интеллекта;

4) обеспечение информационной безопасности;

5) обеспечение конфиденциальности информации;

6) обеспечение непрерывности деятельности.

5.2. В рамках проверки качества наборов данных для разработки искусственного интеллекта организациям рекомендуется проверять используемые наборы данных, в том числе на предмет их достоверности, точности и полноты, собирать и хранить информацию об источниках наборов данных, разметке данных, исправлять неточности в наборах данных и осуществлять их актуализацию.

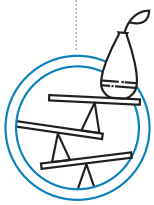
5.3. В рамках проверки качества искусственного интеллекта организациям рекомендуется установить показатели качества искусственного интеллекта и проверять при разработке искусственного интеллекта соответствие искусственного интеллекта установленным организацией показателям качества, в том числе посредством валидации, добровольной сертификации.

5.4. В рамках мониторинга качества искусственного интеллекта организациям рекомендуется регулярно осуществлять проверку применяемого искусственного интеллекта на предмет соответствия установленным организацией показателям качества.

5.5. В рамках обеспечения информационной безопасности организациям рекомендуется учитывать риски искусственного интеллекта, связанные с нарушением информационной безопасности и операционной надежностью, и осуществлять оценку достаточности имеющихся и планируемых к реализации мер противодействия угрозам информационной безопасности на всех этапах разработки и применения искусственного интеллекта, в том числе в отношении наборов данных.

5.6. В рамках обеспечения конфиденциальности информации организациям рекомендуется разработать систему технологических и организационных мер по обеспечению безопасности данных ограниченного доступа, используемых при разработке и применении искусственного интеллекта, в том числе меры, связанные с их обезличиванием и противодействием их несанкционированному распространению при применении больших генеративных моделей сотрудниками организации и ее клиентами.

5.7. В рамках обеспечения непрерывности деятельности организациям рекомендуется обеспечить выполнение значимых функций, осуществляемых с применением искусственного интеллекта, сотрудниками организации в случае временной невозможности применения искусственного интеллекта.



6. ПРИНЦИП ОТВЕТСТВЕННОГО УПРАВЛЕНИЯ РИСКАМИ

6.1. Организациям рекомендуется организовать и осуществлять управление рисками при разработке и применении искусственного интеллекта в рамках общей системы управления рисками в организации.

6.2. В рамках организации управления рисками при разработке и применении искусственного интеллекта организациям рекомендуется с учетом особенностей вида и масштаба ее деятельности, организационной структуры, существующей системы управления рисками и характера услуг, оказываемых организацией, разработать и утвердить правила управления рисками искусственного интеллекта, являющиеся составной частью применяемых документов об управлении рисками в организации, или в виде самостоятельного документа, а также назначить лицо или подразделение, ответственное за соблюдение указанных правил и соответствие их настоящему Кодексу.

6.3. Организациям рекомендуется при управлении рисками искусственного интеллекта обеспечить следующие процессы:

1) учет применяемых моделей искусственного интеллекта;

2) выявление рисков искусственного интеллекта;

3) оценка и присвоение уровня риска искусственного интеллекта;

4) мониторинг и контроль рисков искусственного интеллекта;

5) минимизация выявленных рисков искусственного интеллекта;

6) реагирование на реализовавшиеся риски искусственного интеллекта (далее – риск-события);

7) ведение базы риск-событий.

6.4. В рамках учета применяемых моделей искусственного интеллекта организациям рекомендуется вести централизованный учет моделей искусственного интеллекта и случаев их применения с учетом присвоенного уровня риска искусственного интеллекта.

6.5. В рамках выявления рисков искусственного интеллекта организациям рекомендуется документировать информацию о порядке разработки искусственного интеллекта, включая описание процесса разработки искусственного интеллекта, использованных при разработке наборов данных, сведения об условиях применения искусственного интеллекта, лицах, ответственных за применение искусственного интеллекта в организации, а также анализировать указанную информацию на предмет наличия рисков искусственного интеллекта.

6.6. В рамках оценки и присвоения уровня риска искусственного интеллекта организациям рекомендуется сформировать систему факторов риска и уровней риска, присваивать уровни риска разрабатываемому и применяемому искусственному интеллекту с учетом установленных факторов риска.

6.7. При присвоении уровня риска искусственного интеллекта рекомендуется учитывать следующие факторы риска:

1) сфера применения искусственного интеллекта (например, применение искусственного интеллекта при оказании услуг клиентам, в системах управления рисками, при управлении имуществом организации, в критически важных процессах организации, в процессах, для которых установлены требования операционной надежности);

2) использование при разработке искусственного интеллекта данных ограниченного доступа;

3) размер убытков или ущерб деловой репутации, которые могут быть причинены организации в случае реализации риска;

4) количество клиентов, при оказании услуг которым применяется искусственный интеллект;

5) объяснимость решений, принимаемых искусственным интеллектом;

6) применение искусственного интеллекта, разработанного третьим лицом;

7) использование наборов данных, полученных от третьих лиц, или наборов данных, находящихся в открытом доступе в информационно-телекоммуникационной сети «Интернет»;

8) наличие риск-событий, связанных с применением искусственного интеллекта.

6.8. В рамках мониторинга и контроля рисков искусственного интеллекта организациям рекомендуется регулярно проверять процессы, связанные с разработкой и применением искусственного интеллекта на предмет наличия рисков, а также осуществлять предварительный, текущий, последующий контроль рисков с учетом специфики конкретного случая применения искусственного интеллекта.

6.9. В рамках минимизации выявленных рисков искусственного интеллекта организациям рекомендуется определять меры, направленные на снижение уровня риска искусственного интеллекта, и лиц, ответственных за реализацию указанных мер, а также обеспечивать контроль сотрудников организации за решениями искусственного интеллекта, которому присвоен высокий уровень риска.

6.10. В рамках реагирования на реализовавшиеся риски искусственного интеллекта организациям рекомендуется определять типы риск-событий и принимать меры, направленные на минимизацию последствий риск-событий, а также обеспечить своим сотрудникам и клиентам возможность направления жалоб и сообщений о риск-событиях.

6.11. В рамках ведения базы риск-событий организациям рекомендуется выявлять риск-события и осуществлять их регистрацию с описанием риск-события, обстоятельств, повлекших наступление риск-события, а также приведением сведений о мероприятиях, направленных на минимизацию последствий риск-события, и статусе исполнения таких мероприятий.