

Руководящие принципы в сфере роботов общего назначения

Современные технологические достижения позволяют создавать высокоинтеллектуальных роботов, внешне приближенных к человеку и обладающих высоким уровнем сходства с ним по поведенческим характеристикам (далее – роботы общего назначения). Роботы общего назначения потенциально способны выполнять широкий спектр задач, исторически выполняемых человеком, начиная от опасных работ на производстве, обслуживания клиентов и заканчивая их применением при оказании медицинских услуг.

С каждым годом интерес к индустрии робототехники и, в частности, роботов общего назначения значительно возрастает, что порождает новые вопросы и опасения. Роботы общего назначения и антропоморфные роботы потенциально способны радикально изменить многие отрасли экономики, но их массовое применение поднимает важные этические вопросы, требующие особого внимания. Среди таковых: вопросы конфиденциальности и безопасности, влияние на занятость населения и трудоустройство, уровень автономности роботов и распределение ответственности за их действия.

Эти и другие аспекты должны приниматься во внимание при создании и эксплуатации роботов общего назначения, чтобы гарантировать их этическое применение и благоприятное воздействие на общество и человека.

Настоящие руководящие принципы в сфере роботов общего назначения (далее – Руководящие принципы) разработаны в дополнение и на основе Кодекса этики в сфере ИИ, созданного Альянсом в сфере ИИ, и предлагают всем заинтересованным лицам рекомендации и руководства по этическому подходу к проектированию, созданию и применению роботов общего назначения (далее - РОН) в соответствии с принципами нравственности и этики.

I. Общие положения

1. Область и сфера применения положений Руководящих принципов

1.1. Руководящие принципы являются отраслевым документом, который разработан для конкретизации положений Кодекса этики в сфере ИИ применительно к области РОН в целом и, в частности, к антропоморфным роботам.

1.2. Руководящие принципы предусматривают общие этические принципы и рекомендации, направленные на обеспечение этичности на всех этапах жизненного цикла роботов общего назначения и технологий, связанных с ними.

1.3. Рекомендации настоящих Руководящих принципов рассчитаны на роботов общего назначения, применяемых исключительно в гражданских (невоенных) целях.

1.4. Присоединение к Руководящим принципам лиц, разрабатывающих, внедряющих, использующих и поставляющих на рынок роботов общего назначения (далее – Актеры РОН), является добровольным.

1.5. Положения Руководящих принципов могут быть расширены и/или конкретизированы для отдельных групп Актеров РОН путём составления отдельных дополнений и приложений к Руководящим принципам, разрабатываемым с учетом развития технологий, особенностей решаемых задач, класса и назначения роботов общего назначения, уровня возможных рисков, а также специфического контекста и среды, в которой применяются РОН.

1.6. Положения настоящих Руководящих принципов носят рекомендательный характер.

2. Цели Руководящих принципов

Целью Руководящих принципов является создание условий для:

- 2.1. защиты прав и свобод человека на всех этапах жизненного цикла роботов общего;
- 2.2. обеспечения безопасности и этичности разработки, производства, внедрения, использования и утилизации роботов общего назначения;
- 2.3. предупреждения негативных последствий использования роботов общего назначения;
- 2.4. развития науки и технологий в области робототехники и роботов общего назначения;
- 2.5. поддержки инноваций и конкурентоспособности отечественной робототехнической отрасли, поддержки российских разработчиков роботов общего назначения;
- 2.6. укрепления доверия общества к использованию роботов общего назначения;
- 2.7. повышения информированности и осведомленности общественности о возможностях и ограничениях по использованию роботов;
- 2.8. стимулирования сотрудничества между разработчиками, производителями, пользователями и владельцами роботов;
- 2.9. формирования этических стандартов и принципов, служащих ориентиром для всех участников рынка робототехники;
- 2.10. обеспечения соответствия сценариев и методов использования роботов общего назначения законодательству Российской Федерации и международным нормам.

3. Термины и определения

Термины и определения опираются на ГОСТ Р 60.0.0.4-2019/ИСО 8373:2012 «Роботы и робототехнические устройства. Термины и определения (Robots and robotic devices. Terms and definitions)».

- 3.1. Робот – исполнительный механизм, программируемый по двум или более степеням подвижности, обладающий определенной степенью автономности и способный перемещаться во внешней среде с целью выполнения задач по назначению.
- 3.2. Робот общего назначения (РОН) – многопрофильный универсальный робот, предназначенный для выполнения широкого спектра задач в заданных условиях. РОН обладает автономностью, способностями к передвижению, восприятию окружающей среды и взаимодействию с людьми.
- 3.3. Автономность – способность выполнять задачи по назначению на основе текущего состояния и восприятия внешней среды без вмешательства человека.
- 3.4. Антропоморфный робот – робот с телом, головой и конечностями, выглядящий и перемещающийся подобно человеку.

3.5. Акторы РОН:

- 3.5.1. Разработчики – люди или организации, ответственные за проектирование и разработку роботов общего назначения.
- 3.5.2. Производители – компании, которые производят компоненты или системы для роботов общего назначения.
- 3.5.3. Пользователи – операторы, потребители и организации, которые используют роботов общего назначения для выполнения определенных задач.
 - 3.5.3.1. Оператор – человек, осуществляющий запуск, мониторинг и остановку работ по предназначению робота.
 - 3.5.3.2. Потребитель – лицо, взаимодействующее с роботом с целью получения выгоды.
- 3.5.4. Владельцы – люди или организации, владеющие роботами общего назначения.

II. Принципы этики и рекомендации в сфере роботов общего назначения

4. Человеко-ориентированный подход

4.1. При разработке, внедрении и использовании роботов общего назначения ключевым приоритетом являются права и законные интересы отдельных лиц и общества в целом.

4.2. Взаимодействие с роботами не должно негативно влиять на способность человека к самореализации, на его трудовую и интеллектуальную деятельность. Робот не должен претендовать на те функции и роли, которые человек хочет оставить за собой. Право решать, какую часть своей жизни доверить роботу, всегда остаётся за человеком.

4.3. Необходимо обеспечить равные возможности и доступ к использованию РОН для всех, независимо от расы, пола, возраста, национальности, вероисповедания или социального положения.

4.4. Следует стремиться к обеспечению справедливости и равенства в распределении выгод, связанных с использованием РОН.

4.5. Акторам РОН следует применять этические принципы на всех этапах жизненного цикла роботов общего назначения.

4.6. РОН должны разрабатываться таким образом, чтобы минимизировать появление у людей негативных эмоций (например, страха, как проявления так называемого эффекта «зловещей долины») в результате взаимодействия с РОН.

4.7. Разработчикам рекомендуется предусматривать в работе РОН алгоритмы, соответствующие культурным особенностям и духовно-нравственным ценностям народов Российской Федерации. Следует проектировать РОН таким образом, чтобы РОН транслировали этичное поведение при взаимодействии с человеком.

4.8. Разработка и использование РОН для удовлетворения сексуальных потребностей и/или потребностей в эмоциональной близости человека, а также для создания иллюзии подлинной близости между человеком и РОН, не допускаются.

4.9. Разработчикам и производителям РОН следует предпринимать меры, необходимые для того, чтобы избежать чрезмерной зависимости человека от робота.

5. Непричинение вреда

5.1. Обеспечение безопасности и благополучия человека при взаимодействии с РОН является основополагающим принципом разработки, внедрения и использования РОН.

5.2. Недопустимо использование роботов общего назначения в целях, противоречащих законодательству РФ, включая нарушение прав и законных интересов отдельных лиц, причинение физического, психологического или эмоционального вреда людям, иным живым существам и причинение вреда имуществу, за исключением случаев, предусмотренных в п. 7.2.

5.3. Разработчикам необходимо использовать алгоритмы и предусмотреть меры, не позволяющие РОН совершать противоправные и/или неэтичные действия, которые могут повлечь за собой вред для человека, иного живого существа или имущества.

5.4. Разработчикам и производителям необходимо предусмотреть меры и протоколы безопасности, не позволяющие РОН следовать противоправным или неэтичным инструкциям со стороны пользователя РОН, в частности, способным повлечь за собой вред для человека, иного живого существа или имущества.

6. Роботы общего назначения и право человека на труд

6.1. При внедрении роботов общего назначения необходимо проводить оценку влияния использования РОН на предусмотренное законодательством право человека на труд. Следует проводить оценку рисков нарушения права людей на труд в каждом конкретном случае.

6.2. Внедрение роботов должно осуществляться сбалансировано с учетом прав и интересов работников.

6.3. Работодателям рекомендуется предпринимать разумные и достаточные меры для своевременной переподготовки работников, которые могут быть заменены роботами общего назначения. В качестве лучших практик работодателям рекомендуется оказывать содействие таким работникам в дальнейшем трудоустройстве и/или предоставлять иные меры поддержки, в том числе финансовые.

6.4. Работодателям рекомендуется воздерживаться от злоупотребления автоматизацией рабочих процессов, способной нарушить право работников на труд и/или привести к дискриминационным практикам со стороны работодателя.

6.5. При внедрении РОН работодателям рекомендуется внедрять подходы и формировать культуру коллаборативного взаимодействия сотрудников и роботов.

7. Роботы как инструмент решения задач

7.1. Роботы общего назначения являются инструментами решения рутинных задач для человека, перечень которых может быть определен отраслевыми документами. В этом контексте РОН являются имуществом и собственностью

владельца робота, либо собственностью компании-поставщика услуг при модели поставки Robot-as-a-service (RaaS).

7.2. Акторы РОН должны бережно относиться к роботам при эксплуатации. Недопустимо умышленное причинение вреда роботу, поскольку это потенциально может привести к рискам проекции аналогичного отношения на людей и нарушить права третьих лиц, включая право собственности в отношении РОН. Исключением могут являться случаи, когда действия, способные нанести вред РОН, необходимы для предотвращения и/или прекращения причинения вреда человеку, иному живому существу или имуществу.

8. Подконтрольность

8.1. Технологии роботов общего назначения в любом случае должны находиться под контролем человека, и только люди должны обладать необходимыми полномочиями по принятию итоговых решений в случаях, когда такие решения влекут за собой существенные риски для прав и законных интересов человека, а равно могут повлечь за собой риски нарушения общественных интересов.

8.2. Роботы общего назначения могут обладать определенной степенью автономности. При этом разработчикам следует разрабатывать РОН таким образом, чтобы они всегда действовали исключительно в рамках заранее определенных и законных инструкций (заданий) владельца и/или оператора РОН.

9. Безопасность

9.1. Акторам РОН рекомендуется внедрять комплексные меры информационной безопасности, соответствующие актуальным стандартам в области безопасности, включая внутренние механизмы защиты роботов общего назначения от несанкционированного вмешательства в работу роботов третьими лицами, в том числе посредством вредоносного ПО. Следует информировать пользователей РОН о рисках такого вмешательства и о правилах информационной безопасности при использовании роботов общего назначения.

9.2. Разработчикам и производителям РОН рекомендуется создавать систему и протоколы раннего предупреждения о рисках нарушения безопасности, а также план быстрого реагирования на такие нарушения.

9.3. Разработчикам РОН следует предусмотреть возможность аварийного прерывания работы РОН, в частности, его полного отключения, в случаях утраты эффективного контроля над роботом, а также возникновения угрозы жизни и (или) здоровью человека, риска нанесения ущерба имуществу или наступление иных общественно опасных последствий. Следует предусмотреть меры для минимизации возможных негативных последствий при аварийном отключении РОН, такие как предотвращение резкого падения робота, которое может нанести ущерб людям и имуществу, а также иные возможные негативные последствия.

9.4. Функция отключения РОН может быть реализована в формате автоматической деактивации, ручного отключения или отключения посредством удаленного управления. В любом случае, такой механизм должен быть доступен для пользователя РОН.

10. Прозрачность

10.1. Необходимо обеспечить прозрачность и интерпретируемость алгоритмов искусственного интеллекта, применяемых в роботах общего назначения в степени, необходимой для того, чтобы гарантировать, что действия РОИ будут понятны пользователю. Соответствующую информацию следует доводить до сведения пользователей в объемах, не противоречащих режиму конфиденциальности защищаемой информации.

10.2. Следует предоставлять полную информацию о возможностях и ограничениях РОИ пользователям, что позволит им принимать информированное и осознанно решение относительно использования роботов общего назначения.

10.3. Рекомендуется программировать РОИ таким образом, чтобы они были в состоянии предоставить полную информацию о своих функциях, возможностях и ограничениях по запросу пользователей. РОИ должны уметь объяснять свои действия, намерения и логику принятия решений на понятном для пользователя языке.

10.4. Рекомендуется предусматривать в РОИ подсистемы обработки ошибок для уведомления пользователей о возникающих сбоях, уведомления следует дополнять рекомендациями по устранению неполадок.

10.5. Следует обеспечить пользователям возможность предоставления обратной связи посредством создания каналов для подачи жалоб, направления запросов и внесения предложений по улучшению работы РОИ.

10.6. В целях соблюдения принципа прозрачности при разработке и производстве роботов общего назначения необходимо предусмотреть наличие полной документации, детально описывающей алгоритмы искусственного интеллекта, используемые в таких роботах, данные, которые они собирают и используют, а также другую информацию, касающуюся процесса их создания, производства и эксплуатации. Разработчикам РОИ следует предусмотреть описание сценариев работы РОИ в различных условиях.

11. Конфиденциальность и защита личных данных

11.1. Акторам РОИ следует программировать роботов таким образом, чтобы они обеспечивали конфиденциальность собираемых и обрабатываемых данных.

11.2. Разработка и программирование роботов должны осуществляться в соответствии с принципом *privacy-by-design* (спроектированная защита данных и конфиденциальность по умолчанию) при строгом соблюдении законодательства РФ о защите персональных данных.

11.3. Пользователи должны быть проинформированы о том, какие данные обрабатывает РОИ, где они хранятся и какие меры безопасности применяются для их защиты. Пользовательские данные должны быть защищены по умолчанию.

12. Недискриминация

12.1. Действия РОИ должны быть свободны от предубеждений, предвзятости или стереотипов разработчиков, а равно от дискриминирующих какие-либо группы людей практик и моделей поведения.

12.2. Роботы общего назначения ни при каких обстоятельствах не должны способствовать распространению дискриминационных практик. Напротив, разработчикам и производителям следует учитывать культурные, религиозные и иные

различия при проектировании, разработке и производстве РОН для обеспечения гарантий этичного поведения роботов общего назначения.

13. Ответственность

13.1. Ответственность за вред, наступивший вследствие действий робота общего назначения, возлагается на Акторов РОН согласно общим правилам возникновения юридической ответственности в соответствии с законодательством Российской Федерации.

13.2. Ответственность за действия РОН может быть возложена на любого Актора РОН в зависимости от обстоятельств конкретного дела.

13.3. При этом ответственность Акторов РОН должна соответствовать характеру, степени и размеру вреда, который может возникнуть в результате эксплуатации роботов общего назначения. Для определения степени вины следует учитывать роль Актора РОН в жизненном цикле РОН, его влияние на действия РОН, степень возможного и реального влияния данного Актора РОН на причинение вреда и его размер, а также реальную возможность Актора РОН предотвратить подобный ущерб или вред.

14. Сотрудничество

14.1. Акторы РОН должны следовать приоритету социальной ответственности и устойчивости при разработке и применении технологии роботов общего назначения, способствуя всестороннему развитию этой технологии.

14.2. В целях содействия развитию отрасли роботов общего назначения, стимулирования технологического прогресса, а также минимизации рисков и предотвращения неблагоприятных последствий Акторам РОН рекомендуется:

14.2.1. продвигать международное сотрудничество с разработчиками и производителями РОН в целях обмена знаниями, опытом и технологиями в области робототехники и роботов общего назначения;

14.2.2. наращивать взаимодействие с международным научным сообществом и международными организациями в целях развития и совершенствования технологий РОН, координации усилий в области этики РОН;

14.2.3. сотрудничать с регуляторами и государственными органами Российской Федерации для выработки комплексной политики в области регулирования отрасли РОН;

14.2.4. взаимодействовать с отраслевыми организациями для разработки и внедрения гармонизированных стандартов и рекомендаций в области робототехники, а также создания рейтинга добросовестных производителей и разработчиков РОН;

14.2.5. углублять сотрудничество с общественными организациями и объединениями для учета интересов различных социальных групп при разработке и использовании РОН;

14.2.6. содействовать эффективному взаимодействию между производителями и пользователями РОН для обеспечения качественного обслуживания и поддержки роботов;

14.2.7. укреплять взаимодействие с разработчиками из смежных отраслей и технологий для создания комплексной технологической экосистемы, способствующей инновациям и устойчивому развитию отрасли.

15. Присоединение к Руководящим принципам и его реализация

15.1. Механизм присоединения к настоящим Руководящим принципам и их реализация осуществляется в соответствии с положениями, предусмотренными подразделом 2 раздела 2 Кодекса этики в сфере ИИ.