Научная статья
УДК 34:004:17:296.41:004.8
EDN: https://elibrary.ru/mdiefv
DOI: https://doi.org/10.21202/jdtl.2023.41

Этика искусственного интеллекта и робототехники: ключевые проблемы и современные способы их решения

Нидика Ядав

Независимый исследователь
г. Нью-Дели, Индия

Ключевые слова

ChatGPT,
безопасность,
искусственный интеллект,
кибербезопасность,
конфиденциальность данных,
право,
робот,
робототехника,
цифровые технологии,
этика

Аннотация

Цель: современные достижения в области развития и распространения цифровых технологий привлекли внимание ученых и практиков к обсуждению ключевых этических проблем, связанных с искусственным интеллектом и робототехникой, в связи с чем в настоящем исследовании представлены результаты этой дискуссии и обозначены наиболее актуальные задачи, решение которых определяет пути совершенствования регулирования искусственного интеллекта и робототехники в части морализации технологий.

Методы: в процессе исследования использовались практико- и риск-ориентированный подходы, дополняемые мультидисциплинарным анализом документов (европейских принципов и кодексов этики) и исследований, в том числе посвященных различным проблемам искусственного интеллекта и робототехники.

Результаты: в статье обозначены ключевые этические проблемы в области искусственного интеллекта и робототехники; установлено, что затрагиваемые ключевые этические проблемы могут быть решены при условии, если они получат юридическое оформление и будут реализованы на международном уровне; предложенный автором алгоритм, основанный на анализе практики применения цифровых технологий, позволит усовершенствовать нравственные действия технологий при принятии ими решений.

Научная новизна: в данной статье представлены новейшие этические проблемы, которые волнуют ученых и практиков в области искусственного интеллекта и робототехники, и методы их решения этико-правовыми средствами, направленными на морализацию технологий и повышение ее ответственности.

© Ядав Н., 2023

Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/deed.ru), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Практическая значимость: все решения, представленные в статье, имеют практическое значение и готовы к широкому внедрению на международном уровне. Их оформление в нормативном виде и последующее соблюдение приведут к уменьшению вреда, который может нанести искусственный интеллект в прикладных областях, в том числе в робототехнике с использованием искусственного интеллекта. В связи с этим нормативные, в том числе законодательные, решения должны быть приняты как можно скорее, чтобы искусственный интеллект и робототехника приобрели статус надежных инструментов при использовании этих систем на работе, дома и в других сферах, таких как торговые центры, магазины, школы, университеты и т. д.

Для цитирования

Ядав, Н. (2023). Этика искусственного интеллекта и робототехники: ключевые проблемы и современные способы их решения. *Journal of Digital Technologies and Law*, *1*(4), 955–972. https://doi.org/10.21202/jdtl.2023.41

Содержание

Введение
1. Общая информация об этике искусственного интеллекта и робототехники
2. Робототехника
   2.1. Ответственная робототехника
3. Решения в области этики искусственного интеллекта и робототехники
   3.1. Защита частной жизни
   3.2. Право собственности
      3.2.1. Традиционное искусство
      3.2.2. Искусство с применением технологий и искусственного интеллекта
      3.2.3. Решения
   3.3. Пути к достижению ответственной робототехники
      3.3.1. Беспилотные автомобили на основе искусственного интеллекта
      3.3.2. Искусственный интеллект, робототехника и проблемы занятости
      3.3.3. Языковые модели на основе искусственного интеллекта
Заключение
Список литературы

Введение

Технология искусственного интеллекта (далее – ИИ) (Rich et al., 2009) в последнее время набирает все большие обороты. За последние несколько лет достигнут колоссальный прогресс в каждой из областей применения ИИ (Saveliev et al., 2021). Более того, многие аспекты последних разработок в области ИИ нельзя было и представить (Xiao-Fan et al., 2023; Kumar et al., 2023). Так, даже сами создатели такого ИИ-продукта, как ChatGPT, были ошеломлены тем, каких вершин развития они достигли.

Это лишь один из примеров; прогресс наблюдается не только в генерации текстов, но и во многих областях исследований ИИ. Некоторые из них мы рассмотрим в данной статье.

Термин «искусственный интеллект» был введен для обозначения технологии, которая имитирует работу человеческого мозга для выполнения сложных человеческих действий. При этом преследуется цель облегчить труд человеческий, производя его так же, как это делает человек, или оказывать помощь в решении задач, поставленных перед ним. Предмет ИИ (Rich et al., 2009; Stahl, 2021; Mueller, 2012) представляет собой объединение многих областей, таких как информатика, математика, когнитивные науки, психология и др. ИИ зародился в начале 50-х годов прошлого века и с тех пор привлекает внимание исследователей (Turing, 2009; Schank, 1991). Эта тема разрабатывалась с целью создания алгоритмов и/или машин, которые могли бы выполнять работу или действовать подобно человеку, а в некоторых случаях, когда человек не мог выполнять работу, интеллектуальные машины на основе ИИ могли бы заменить его, например при добыче минералов в земной коре (Lesandrini et al., 2023; Memarian & Doleck, 2023).

Перед ИИ никогда не ставилась цель причинить вред людям, заменить человека или захватить мир людей. Цель ИИ – дополнить человека, помочь людям совершать научные открытия или бороться с природными катаклизмами, помочь человеку в таких ситуациях, как пандемия и т. д. Другие области применения, в которых ИИ, как представляется, может быть полезным людям, – это повседневная медицинская помощь, работа в машиностроительных компаниях, управление полуавтоматическими и полностью автоматизированными автомобилями, доставка грузов с помощью ботов в чрезвычайных ситуациях, например, при наводнениях. Интеллектуальная робототехника может найти применение в самых разных областях, от текстильной промышленности до управления самолетами, от организации дорожного движения до космических программ и т. д.

В настоящее время развитие искусственного интеллекта достигло больших высот, и он может выполнять работу, которую делает человек, практически во всех областях. Искусственный интеллект заменяет множество рабочих мест, но это не должно пугать людей. Новые рабочие места появляются с той же скоростью, однако статистика создания новых рабочих мест на месте исчезнувших не сразу становится известной простым людям. Именно так шло развитие со времен промышленной революции. Да, освоить новую сферу деятельности из-за того, что искусственный интеллект отменил многие рабочие процессы, не так-то просто. Все мы помним, как в результате компьютеризации текстильного производства было ликвидировано множество рабочих мест. Пострадавшим инженерам пришлось искать новую работу. Как только какая-то область устаревает, начинается развитие новой области. Так, с появлением мобильных телефонов стали неактуальными стационарные телефоны. Сегодня те же мобильные телефоны обеспечивают работой множество людей по всему миру. Подробнее о решении проблемы потери рабочих мест будет рассказано в следующих разделах.

Цель данной статьи – рассказать об этических аспектах искусственного интеллекта, его этическом использовании и предложить пути решения проблем, возникающих в связи с развитием ИИ и особенно комбинации ИИ с робототехникой. В разделе 2 описаны причины, побудившие к изучению этических проблем ИИ и робототехники. В разделе 3 рассматриваются вопросы использования ИИ в робототехнике. В разделе 4 предложены решения основных этических проблем в области ИИ и робототехники, а в разделе 5 даны выводы по работе.

1. Общая информация об этике искусственного интеллекта и робототехники

Несмотря на множество полезных применений искусственного интеллекта, многие воспринимают его как угрозу. Однако, как и в других областях науки, ИИ представляет самостоятельный предмет изучения. В данной статье рассматривается необходимость установления этических норм в ИИ и робототехнике. Следует знать, что, как и в случае других наук, необходимо иметь представление об определенных ограничениях при разработке, открытии и внедрении искусственного интеллекта. В этом нет ничего нового, такие хорошо развитые дисциплины, как химия, физика, биология и другие, имеют свои правила использования. Примером может служить утечка выращенных в лаборатории грибков, которые могут создать хаос в природе. Это может произойти в результате отсутствия должного внимания и несоблюдения протоколов лаборатории, в которой грибки хранились для проведения практических работ. Таким образом, во всех областях науки существуют меры защиты; они должны применяться и в области искусственного интеллекта и робототехники. Другой пример – новый автомобиль, в котором для работы двигателей используется ядерная энергия. Если его не протестировать в экстремальных температурных условиях, это может привести к многочисленным проблемам.

Разница между другими научными дисциплинами и ИИ заключается в том, что ИИ вырос из ряда научных дисциплин, в основном информатики и математики, в которых не существует протоколов выпуска результатов. Таким образом, выпуск продуктов ИИ начался без научных и социальных рекомендаций по их внедрению. Другая причина заключается в том, что ИИ все еще является новшеством по сравнению с другими дисциплинами, и это необходимо учитывать при определении ограничений на производство устройств с искусственным интеллектом.

К счастью, серьезных регулярных сбоев в работе искусственного интеллекта и робототехники пока не наблюдается. Некоторое время назад возник ряд проблем, касавшихся автономных автомобилей и системы рекомендательных норм одного из технологических гигантов. В будущем появится еще больше продуктов ИИ и робототехники, но они должны быть этичными, поскольку на них лежит большая ответственность. ИИ помог человеку достичь того уровня развития, который существует сегодня. В той или иной мере искусственным интеллектом оснащены многие устройства и продукты, которыми мы пользуемся, от стиральной и посудомоечной машины до онлайн-бронирования билетов на отдых. Все это будет использоваться и в будущем. Развитие ИИ невозможно остановить, поэтому для более полного раскрытия потенциала ИИ и робототехники необходимо соблюдать этические нормы.

Сейчас ИИ является самостоятельным объектом. Необходимо выработать специальные инструкции для ИИ. Как физические эксперименты проводятся согласно заранее определенным инструкциям, так и все эксперименты в области ИИ и робототехники должны следовать установленным инструкциям. Это должно быть сделано для каждого приложения ИИ. Составление инструкций может занять некоторое время, но это необходимо, поскольку правила будут определяться на научной основе. Соответствие правилам должно проверяться до выпуска на рынок каждого продукта ИИ, будь то робототехника или простая программа (Rich et al., 2009), запускающая песню, исходя из ваших предпочтений. Таким образом, необходимо установить соответствующие правила для всех разработчиков ИИ.

2. Робототехника

Понятие «робототехника» очень обширное и до сих пор достаточно размытое. Как правило, роботы определяются как устройства, которые способны совершать механические движения и могут быть как с искусственным интеллектом, так и без него. Соответственно, робототехника – это дисциплина, изучающая роботов (Agarwal & Stoff, 2023). Однако в данной статье речь пойдет об интеллектуальных роботах (Joachim et al., 2021), которые могут управляться и работать на основе искусственного интеллекта. В сфере робототехники применяются как простые механические роботы, так и роботы для медицинской хирургии, и даже гуманоидные роботы, способные к самостоятельным исследованиям и полетам в космос, но при этом так или иначе управляемые человеком.

Есть и другие проблемы с определением понятия «робот» как машины с дистанционным управлением для выполнения тех или иных задач. По этому поводу также ведутся дискуссии (De Felice et al., 2022), называть ли такое устройство интеллектуальным роботом. На этот вопрос отвечают по-разному, но мы считаем, что такая машина является интеллектуальным роботом, поскольку она управляется не человеческим мозгом, а сигналами, посылаемыми на расстоянии через Интернет или другие источники. Примером таких устройств являются, например, боты на базе интернета вещей. Тем самым определение робототехники, причем именно интеллектуальной робототехники, расширяется. Они имитируют работу мозга, хотя и посредством канала связи. Таким образом, законы этики в робототехнике распространяются и на все подобные устройства.

Почему анализ робототехники важен при обсуждении этики ИИ? Причина этого в том, что робототехника – это будущее, к которому движется ИИ. Роботы всегда должны контролироваться и изменяться людьми, под руководством которых выполняется поставленная задача. Здесь возникает этика для роботов, использующих искусственный интеллект. Этика робототехники касается не только управления выполняемой работой, но и влияния роботов с искусственным интеллектом на окружающих их людей, находящихся в зоне выполнения задачи.

2.1. Ответственная робототехника

Ответственная робототехника (Memarian & Doleck, 2023) – это робототехника, целью которой является помощь человеку в выполнении различных задач: от повседневных домашних дел до сложных операций и даже полетов в космос для совершения научных открытий. Однако, являясь основной частью ИИ, робототехника не должна наносить вред человеку и окружающей его среде. Именно это должно стать целью создания ответственной робототехники, именно это должно лежать в основе создания интеллектуальной робототехники на базе ИИ. Следует отметить, что ответственная робототехника не может пойти по неверному пути, поскольку входные данные преобразуются в выходные. Для этого необходимо разработать этические протоколы.

Необходимо ответить на важный вопрос: зачем нам нужны роботы? Они нужны для того, чтобы помогать людям либо выполнять то, что люди делать не могут. Роботы работают на языках программирования, используя цифровые данные. Интеллектуальная робототехника должна понимать проблемы, прежде чем она сможет решать их.

В настоящее время на производстве применяются различные виды роботов, в том числе с искусственным интеллектом (Brooks, 1991). Их можно использовать в спасательных операциях на шахтах, для колонизации других планет или космических объектов. В целях безопасности робототехника должна быть отправлена в космическое пространство раньше самих людей. Пока не найдена другая планета, похожая на Землю, но человечество живет надеждой, и с надеждой люди достигли невероятного прогресса во всех отраслях. Можно попытаться создать такой ИИ и роботов, которые помогут сохранить жизнь и благополучие на Земле, и одновременно создать условия для жизни на других космических телах.

Роботы на основе ИИ могут помочь и при экстремальных лесных пожарах. Они могут извлекать из мусора металлы и другие ценные вещества, автоматически отделять одноразовый пластик от других видов мусора. Есть работа, не предназначенная для человека, и тогда на помощь приходят машины и роботы, основанные на искусственном интеллекте. Вместе с тем появляется и ответственность человека за то, чтобы сделать эти машины безопасными для использования. Еще раз подчеркнем, что ИИ находится на стадии зарождения и раннего развития, поэтому необходимо соблюдать этические нормы во всех случаях применения ИИ.

3. Решения в области этики искусственного интеллекта и робототехники

Если задана программа и известно, что машина всегда следует заложенным в нее правилам, то можно быть уверенным, что ответственная робототехника не может пойти по ложному пути. Существует две характеристики ИИ с точки зрения его использования и получаемых результатов. Они представлены следующим образом (Kumar et al., 2023):

1. Хороший ИИ.

Хороший ИИ – это полезный ИИ, разработанный для выполнения задач и проверяемый с помощью методов «черного ящика» и «белого ящика». Такой ИИ требует определенных входных данных и выдает предсказуемые результаты. Он может использовать обычные инструменты распознавания образов (Duda & Hart, 2006); к примеру, с помощью метода опорных векторов на основе дерева решений прогнозируется вероятность возникновения лейкемии. Эти алгоритмы не могут причинить никому вреда и обычно имеют заранее заданные выходные параметры. Другие области применения такого ИИ – оборона, полеты самолетов, запуск ракет, прогнозирование состояния здоровья, медицинские операции. Таким образом, эти алгоритмы можно назвать хорошим ИИ.

2. Плохой ИИ.

Алгоритм ИИ может нанести вред, используя необъективные данные, нарушая конфиденциальность, и по многим другим причинам. Проблемы могут вызвать различные приложения для обработки изображений, редактирования видео, создания поддельных профилей и фейковых новостей, взлома, кражи интеллектуальной собственности и т. д. Все это может причинить ущерб личной и профессиональной жизни людей.

Все основные проблемы, требующие решения, являются результатом использования плохого ИИ. Хороший ИИ отвечает требованиям этики и не наносит вреда. В нем, как правило, используются алгоритмы машинного обучения (Bishop, 2006; Palladino, 2023) для поиска решений и работы над проблемами. И все же для большей безопасности должны быть разработаны рекомендации для всех видов ИИ – хорошего и плохого, простого ИИ или AGI (general intelligence-based AI) – ИИ, основанного на общем интеллекте. Для любого из этих видов ИИ и робототехники с ИИ необходимо проводить проверки как по методу «черного ящика», так и по методу «белого ящика». Эти проверки позволяют в определенной степени предотвратить ошибочные результаты. Тестирование по методу «черного ящика» может выполнить любой человек, владеющий компьютерной грамотностью, в то время как тестирование по методу «белого ящика» – более комплексное, оно предполагает работу с кодом для выяснения причин возникшей ошибки. Таким образом, проблема решается в корне. Тестирование по методу «белого ящика» должно широко использоваться в отношении этического ИИ в робототехнике и интернете вещей.

3.1. Защита частной жизни

Такие инструменты на основе искусственного интеллекта, как «помощники по созданию текстов» и «умный ответ», могут вызывать опасения по поводу конфиденциальности (Zhang et al., 2021), поскольку создается впечатление, что ваши электронные письма читаются искусственным интеллектом и могут быть использованы не по назначению. Проблемы защиты частной жизни не ограничиваются электронной почтой. Они могут распространяться на конфиденциальность данных, когда существует риск утечки информации, и на личную конфиденциальность, когда под угрозой оказывается профиль пользователя аппаратного/программного обеспечения ИИ. Это могут быть данные о личности пользователя, о его компании, семье или имуществе. Кроме того, любое посягательство на частную жизнь является преступлением, поэтому инструменты и приложения на основе ИИ должны быть защищены от атак на неприкосновенность частной жизни или вирусных угроз. Приложения на основе искусственного интеллекта должны быть сделаны таким образом, чтобы избежать атак вирусов. Кроме того, сами приложения ИИ могут генерировать данные о просмотре веб-страниц и других действиях пользователя; обо всем этом следует позаботиться до начала использования инструмента ИИ.

Решением этой проблемы являются анализ собственных действий и установка соответствующих патчей на системы, использующие ИИ, то есть установка защиты от вредоносного ПО и вирусов. Для других приложений, использующих автоматическое наполнение на основе искусственного интеллекта, необходимо требовать от производителя соблюдения установленных норм для инструментов искусственного интеллекта. Только так можно обеспечить безопасность наших данных.

3.2. Право собственности

С развитием искусственного интеллекта возникает проблема прав собственности. Если изначально ИИ помогал художникам, авторам песен, сценаристам и создателям фильмов, то теперь он захватил все эти области. Кто научил ИИ всему этому? Ответ – мы сами. Программы на основе искусственного интеллекта были обучены

на сотнях и тысячах созданных людьми произведений искусства, фотографий, чатов и текстов. Когда ИИ генерирует, скажем, новое изображение, то он использует тысячи произведенных человеком предметов, а значит, заслуга искусственного интеллекта в создании новых изображений принадлежит и людям, которые не только создали исходное изображение, но и научили ИИ рисовать или, к примеру, петь.

Решение этой проблемы можно представить следующим образом. Существуют традиционные художники и художники на основе ИИ и цифровых искусств. Понятие искусства должно быть ограничено тем, что создано живописцами; аналогично следует рассуждать и в отношении других творцов, будь то сценаристы, музыканты и т. д.

3.2.1. Традиционное искусство

Традиционные художники придерживаются как освященных веками, так и новых подходов к созданию произведений искусства. Многие художники создают произведения в традиционных цветах и стилях, кто-то выбирает акриловые краски, а кто-то старое доброе масло. Другие работают со стеклом или глиной. Все это – примеры традиционных способов создания произведений искусства. Для них искусство – не только профессия, но и средство к существованию, страсть, источник любви и предмет поклонения. Большинство художников пишут не только для души, но и для получения побочного дохода, поэтому они могут пострадать не так сильно, как те, кто полностью полагается на творчество.

С развитием искусственного интеллекта, появлением новых потребностей и технологий старые способы создания произведений искусства могут исчезнуть. Новые технологии позволяют создавать произведения искусства быстрее, как правило, дешевле и с большей прибылью (Scimeca & Bonfiglio, 2023).

Однако большинство людей все же предпочитают традиционные произведения искусства. Привычные способы их создания требуют больше усилий, чем сравнимое по качеству цифровое искусство. Для многих освоение нового направления является непростой задачей, поэтому они придерживаются старых способов создания произведений искусства. Поначалу цифровое искусство также нелегко освоить, но когда художник овладеет им, то его развитие не потребует больших усилий.

3.2.2. Искусство с применением технологий и искусственного интеллекта

Художники, работающие в направлении техно и ИИ, используют те или иные технологии или искусственный интеллект для создания произведений искусства любого рода, будь то музыка, сценарии или картины. Они могут использовать такие инструменты, как программное обеспечение, 2D- или 3D-печать, искусственный интеллект и роботизированные руки. Изучив этот инструментарий, они могут погрузиться в безграничный мир творчества. Инструменты, управляемые искусственным интеллектом, еще более интересны, хотя для создания шедевров в музыке и живописи может потребоваться написать несколько строк кода. Здесь художник может работать как в сотрудничестве, так и самостоятельно, как создавать произведения с нуля, так и модифицировать и улучшать чужие работы.

Работы могут быть созданы на основе произведений, которые не принадлежат художнику, создающему новую работу. В таком случае юридически владелец оригинального произведения искусства обладает авторскими правами на него.

До сих пор не существует четкого законодательства, регулирующего эту сферу. В этом случае могут пострадать обе стороны: и цифровой художник, поскольку он вложил свои навыки работы с искусственным интеллектом в чужое произведение, и автор оригинального произведения, не получивший заслуженного вознаграждения. Пока правовые вопросы в этой сфере не решены, преимущество остается за автором оригинального произведения.

В данной ситуации как традиционные художники, так и те, кто работает с цифровыми технологиями и искусственным интеллектом испытывают огромное давление. Проблемы возникают в отношении авторских прав, лицензий на использование технологий, вопросов оплаты, репутации художника, прав собственности на произведения искусства, подтверждения их подлинности.

3.2.3. Решения

Искусственный интеллект и робототехника обучаются на тысячах созданных человеком произведений искусства, фотографий, сценариев и т. д. Эти обучающие данные используются для создания новых произведений искусства. Однако не всегда понятно, как найти создателей тех произведений искусства, на основе которых искусственный интеллект сгенерировал свое произведение. Иногда вклад заключается лишь в использовании одного оттенка цвета, а в другом случае используется целиком лицо, обработанное фильтром. Необходимо определить, что было взято из исходных обучающих данных, а что создано самим искусственным интеллектом. Юридически решение можно сформулировать следующим образом:

1. Если художественное произведение, созданное с помощью искусственного интеллекта, имеет более 80 % сходства с оригинальным произведением в обучающих данных, то создатель оригинального произведения и цифровой художник могут договориться о разделе прибыли или запрете на продажу нового произведения по желанию создателя оригинального произведения. Последний может купить работу, созданную цифровым художником, исходя из количества часов, потраченных им на редактирование исходного произведения.

2. В случае если доля оригинала составляет менее 80 % цифрового произведения, создатель оригинала может претендовать на часть прибыли от продажи работы и не может запретить цифровому художнику продавать ее.

3. Если создатель оригинального произведения пожелает работать вместе с цифровым художником, то они смогут разделить прибыль; такое сотрудничество может привести к новым высотам в создании невиданных ранее произведений искусства.

Такие решения могут гармонизировать рынок прав собственности на произведения искусства.

3.3. Пути к достижению ответственной робототехники

Среди проблем ответственной робототехники нужно выделить следующие (Stahl et al., 2023):

1. Конкурирующие роботы. По мере развития робототехники с искусственным интеллектом могут возникнуть такие проблемы, как противостояние роботов, созданных в различных странах. Эти контакты должны быть мирными; роботы не должны бороться друг с другом, они должны работать совместно, не создавая глобальной напряженности и конфликта интересов.

2. Полностью автономные роботы. Эти роботы, обладающие самосознанием, имеющие собственные мыслительные процессы и способные самостоятельно принимать решения, будут созданы на основе общего искусственного интеллекта (Artificial General Intelligence, AGI). Они находятся в стадии создания, и, насколько нам известно, ни один робот на базе AGI еще не используется. На таких роботов должны быть наложены особые ограничения.

3. Сбои в работе роботов на основе ИИ. Подобные проблемы могут представлять реальную угрозу для нормального функционирования того процесса, для которого был создан робот. Единственный способ решения этой проблемы – вывод робота из строя и установка программ, необходимых для устранения неполадок. После этого проводят соответствующее тестирование, а затем того же или нового робота вновь запускают в работу.

Решение заключается в создании универсальных машин, не конкурирующих между собой для достижения результатов и работающих в соответствии с этикой благополучия человека. Для этого необходимо на глобальном уровне договориться о нейтральности ИИ и робототехники, принятии и соблюдении законов их использования для достижения взаимовыгодных результатов. Таким образом, страны мира должны подписать юридические документы, гарантирующие сотрудничество как в области ИИ, так и в области робототехники. Договоренность должна предусматривать, что робототехнические устройства, создаваемые странами в оборонном, государственном, общественном и частном секторах, могут использоваться только с одобрения всего мирового сообщества.

3.3.1. Беспилотные автомобили на основе искусственного интеллекта

Беспилотный автомобиль с искусственным интеллектом должен не только доехать из точки А в точку Б, но и соблюдать правила дорожного движения, ездить по соответствующим дорогам, правильно парковаться, соблюдать скоростной режим, не наезжать на людей и животных по пути следования. Лишь при соблюдении этих условий беспилотный автомобиль может быть выпущен на дорогу. Недопустимы ситуации, подобные той, которая произошла несколько лет назад, когда беспилотный автомобиль столкнулся на дороге с другим автомобилем, определив это как ошибку. Согласно этике ИИ, такие автомобили не должны использоваться на дорогах, пока не будут пройдены все обязательные проверки на этику ИИ.

Следует отметить, что какого-либо одного закона в области робототехники будет недостаточно. Рассмотрим проблему этики беспилотных автомобилей на примере общей этики робототехники. Каждая область применения робототехники отличается от другой, поэтому сначала необходимо дать определения, затем указать, какую роль должны выполнять роботы на основе искусственного интеллекта. Затем юристы определяют, какие организации и субъекты могут быть затронуты процессами с участием робота на основе искусственного интеллекта; для каждого отдела разрабатывается схема безопасной работы. Отдел по этике совместно с техническим отделом по применению роботов должен разработать рекомендации по безопасному использованию роботов. Когда компания, создающая роботов, будет готова к их внедрению, отдел этических стандартов ИИ проводит проверку всего процесса. Лишь после получения одобрения всех сторон роботы могут быть допущены к работе (Leeson & Coyne, 2005).

3.3.2. Искусственный интеллект, робототехника и проблемы занятости

С появлением искусственного интеллекта происходит автоматизация существующих областей, что может привести к исчезновению некоторых рабочих мест. Возможными выходами из этой ситуации являются обучение и переобучение. Но такая возможность не гарантируется во всех областях, а также недоступна для некоторых людей по причине их возраста или инвалидности. В таких случаях работодатель должен тщательно подойти к вопросу увольнений. Это этические проблемы, которые приходится решать в качестве побочного эффекта от того, что ИИ выполняет человеческую работу (Duan, et al., 2019). У человека есть огромный потенциал, который он может раскрыть, существует множество новых областей, которые можно исследовать, имея государственное или частное финансирование. Однако переход из одной области в другую должен быть плавным, а компании должны помогать сотрудникам найти новую работу, соответствующую их способностям, заслугам, целям и размеру вознаграждения. Кроме того, должна быть предоставлена гарантия, скажем, на пять лет работы. Уязвимые категории населения должны сохранить свои рабочие места, их нельзя увольнять, как это происходит в 2023 году во всем мире.

Существует множество приложений, которые могут выполнять работы на основе искусственного интеллекта и робототехники. Это не означает, что необходимо провести сокращения рабочих мест и увольнения. Роботы не должны отнимать рабочие места, не создавая новых взамен. Как оборудование места для нового питомца в доме или установка новой люстры не отнимают у вас время для сна, так же должно обстоять дело и с искусственным интеллектом и робототехникой. Это единственный путь вперед – принять изменения, которые принесли с собой ИИ и робототехника, и ожидать появления новых рабочих мест. В то же время все нужно делать осторожно, учитывая ограничения, налагаемые этикой искусственного интеллекта (Hellwig et al., 2019).

3.3.3. Языковые модели на основе искусственного интеллекта

Искусственный интеллект находит широкое применение в больших языковых моделях (Large Language Models, LLM). В 2022 году наблюдался невиданный ранее рост LLM на основе искусственного интеллекта (Ferreira & Lipoff, 2023, Glaser, et al., 2019). То одна, то другая модель оказывались в центре внимания и били рекорды по производительности и точности решения задач, а некоторые по обоим параметрам сразу. При использовании LLM возникают такие проблемы, как непредсказуемость результатов, предвзятость, проблемы конфиденциальности, ошибки при передаче содержания, языковые ошибки и многие другие (Díaz-Rodríguez et al., 2023). Для дальнейшего развития LLM необходимо проводить оценку программного обеспечения искусственного интеллекта по методу «белого ящика», о чем говорилось в начале данного раздела. Проверка по этому методу позволяет предсказать следующий результат при заданной комбинации входных данных. Это должно свести к минимуму такие проблемы, как те, что возникли два года назад у одного из технологических гигантов в связи с использованием ошибочной рекомендательной системы, от которой пострадали многие группы людей.

Заключение

После установки и проверки программного обеспечения машина всегда следует правилам, заложенным в ее программу, и можно быть уверенным, что ответственная робототехника с искусственным интеллектом не пойдет по ложному пути. В данной

работе мы изучили ряд этических вопросов, распространенных в современном мире, и предложили новые решения этих проблем. Кроме того, было рассмотрено сочетание робототехники с искусственным интеллектом с точки зрения его потенциального использования и этики применения в будущем. Решения для многих других этических проблем еще предстоит разработать, что должно стать темой дальнейших исследований.

Список литературы

Agarwal, A., & Stoff, B. (2023). Ethics of using generative pretrained transformer and artificial intelligence systems for patient prior authorizations. *Journal of the American Academy of Dermatology*, *20*, 23–34. https://doi.org/10.1016/j.jaad.2023.04.030

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer. https://clck.ru/37AU6Z

Brooks, R. A. (1991). New approaches to robotics. *Science*, *253*(5025), 1227–1232.

Critchlow, A. J. (1985). *Introduction to robotics*. Macmillan Pub Co.

De Felice, F., Petrillo, A., De Luca, C., & Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. *Procedia Computer Science*, *200*, 1846–1856. https://doi.org/10.1016/j.procs.2022.01.385

Hellwig, J., Huggett, S., & Siebert, M. (2019). *Artificial Intelligence: How knowledge is created, transferred, and used*. Elsevier. https://doi.org/10.17632/7ydfs62gd6.2

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges, and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

Ferreira, A. L., & Lipoff, J. B. (2023). The complex ethics of applying ChatGPT and language model artificial intelligence in dermatology. *Journal of the American Academy of Dermatology*, *89*(4), e157-e158. https://doi.org/10.1016/j.jaad.2023.05.054

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137. https://doi.org/10.1016/j.pneurobio.2019.01.008

Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg, & Marcelo Sánchez Sorond. (2021). AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy. In J. von Braun et al. (Eds.), *Robotics, AI, and Humanity*. https://doi.org/10.1007/978-3-030-54173-6_1

Kumar, P., Chauhan, S., & Kumar, L. A. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. *Engineering Applications of Artificial Intelligence*, *120*, 105894. https://doi.org/10.1016/j.engappai.2023.105894

Leeson, P. T., & Coyne, C. J. (2005). The economics of computer hacking. *JL Econ. & Pol'y*, *1*, 511.

Lesandrini, J., Idris, M. Y., & Reis, D. S. (2023). The Ethics of Artificial Intelligence and Machine Learning. *Journal of Radiology Nursing*, *42*(3), 265–266. https://doi.org/10.1016/j.jradnu.2023.05.001

Memarian, D., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152, https://doi.org/10.1016/j.caeai.2023.100152

Mueller, V. C. (2012). Introduction: philosophy and theory of artificial intelligence. *Minds and Machines*, *22*(2), 67–69. https://doi.org/10.1007/s11023-012-9278-y

Palladino, N. (2023). A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy*, *47*(5), 102479. https://doi.org/10.1016/j.telpol.2022.102479

Rich, E., Knight, K., & Nair, S. (2009). *Artificial Intelligence*. Tata McGraw Hill.

Saveliev, A. M., Zhurenkov, D. A., Poikin, A. E., & Berkutova, T. A. (2021). Ethics of Artificial Intelligence and Post-non-classical Scientific Rationality. *IFAC-PapersOnLine*, *54*(13), 397–401. https://doi.org/10.1016/j.ifacol.2021.10.480

Schank, R. C. (1991). Where's the AI? *AI magazine*, *12*(4), 38. https://doi.org/10.1609/aimag.v12i4.917

Scimeca, M., & Bonfiglio, R. (2023). Comment on redefining authorship in the era of artificial intelligence: balancing ethics, transparency, and progress. *ESMO Open*, *8*(5), 101635. https://doi.org/10.1016/j.esmoop.2023.101635

Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, *16*(3), 26–33. https://doi.org/10.1109/msp.2018.2701164

Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study. *Technological Forecasting and Social Change*, *191*, 122502. https://doi.org/10.1016/j.techfore.2023.122502

Stahl, B. C. (2021). Perspectives on Artificial Intelligence. In *Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance* (pp. 7–17). Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9_2

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23–65). Springer Netherlands.

Xiao-Fan, L., Zhaoyang, W., Wei, Zh., Guoyu, L., & Gwo-Jen, H. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*, *201*, 104813. https://doi.org/10.1016/j.compedu.2023.104813

Zhang, Y., Wu, M., Yijun Tian, G., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994. https://doi.org/10.1016/j.knosys.2021.106994

Информация об авторе

Ядав Нидика – PhD, независимый исследователь
Адрес: 110074, Индия, г. Дели, Чаттапур Анклав Фейз II
E-mail: yadavnidhika68@gmail.com
ORCID ID: https://orcid.org/0000-0002-0165-6453
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57195398958
WoS ID: https://www.webofscience.com/wos/author/record/2269009
Google Scholar ID: https://scholar.google.com/citations?user=fndffSwAAAAJ

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов

Финансирование

Исследование не имело спонсорской поддержки

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.41.51 / Охрана авторских прав
Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 23 июля 2023 г.
Дата одобрения после рецензирования – 25 октября 2023 г.
Дата принятия к опубликованию – 30 ноября 2023 г.
Дата онлайн-размещения – 15 декабря 2023 г.

Research article
DOI: https://doi.org/10.21202/jdtl.2023.41

Ethics of Artificial Intelligence and Robotics: Key Issues and Modern Ways to Solve Them

Nidhika Yadav

Independent Researcher
New Delhi, India

Keywords

artificial intelligence,
ChatGPT,
cyber security,
data privacy,
digital technologies,
ethics,
law,
robot,
robotics,
safety

Abstract

Objective: modern achievements in the development and dissemination of digital technologies have attracted the attention of scholars and practitioners to the discussion of key ethical issues related to artificial intelligence and robotics. Hence, this study presents the most relevant of these issues, posing new challenges for legal scholars and practitioners to develop the regulation of artificial intelligence and robotics in terms of technology moralization.

Methods: the research used practice- and risk-oriented approaches, complemented by multidisciplinary analysis of documents (European principles and codes of ethics) and studies, including those devoted to various problems of artificial intelligence and robotics.

Results: the article identifies key ethical issues in the field of artificial intelligence and robotics. It is established that the key ethical issues involved can be solved if they are legally formalized and implemented at the international level. The algorithm proposed by the author, based on the analysis of the digital technologies application, will allow improving the moral actions of technologies in the process of their decision making.

Scientific novelty: the article presents the latest ethical problems that concern scientists and practitioners in the field of artificial intelligence and robotics, and the methods of their solution by ethical and legal means aimed at moralizing technology and increasing its responsibility.

© Yadav N., 2023

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Practical significance: all solutions presented in the article have practical significance and are ready for wide implementation at the international level. Their formalization in normative form and subsequent compliance will reduce the harm that artificial intelligence may cause in applied fields, including robotics using artificial intelligence. Regulatory, including legislative, decisions must therefore be taken as soon as possible to ensure that artificial intelligence and robotics become reliable tools for these systems to be used at work, at home, and in other areas such as shopping centers, stores, schools, universities, etc.

For citation

Yadav, N. (2023). Ethics of artificial intelligence and robotics: key issues and modern ways to solve them. *Journal of Digital Technologies and Law, 1*(4), 955–972. https://doi.org/10.21202/jdtl.2023.41

References

Agarwal, A., & Stoff, B. (2023). Ethics of using generative pretrained transformer and artificial intelligence systems for patient prior authorizations. *Journal of the American Academy of Dermatology*, *20*, 23–34. https://doi.org/10.1016/j.jaad.2023.04.030

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer. https://clck.ru/37AU6Z

Brooks, R. A. (1991). New approaches to robotics. *Science*, *253*(5025), 1227–1232.

Critchlow, A. J. (1985). *Introduction to robotics*. Macmillan Pub Co.

De Felice, F., Petrillo, A., De Luca, C., & Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. *Procedia Computer Science*, *200*, 1846–1856. https://doi.org/10.1016/j.procs.2022.01.385

Hellwig, J., Huggett, S., & Siebert, M. (2019). *Artificial Intelligence: How knowledge is created, transferred, and used*. Elsevier. https://doi.org/10.17632/7ydfs62gd6.2

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges, and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Duda, R. O., & Hart, P. E. (2006). *Pattern classification*. John Wiley & Sons.

Ferreira, A. L., & Lipoff, J. B. (2023). The complex ethics of applying ChatGPT and language model artificial intelligence in dermatology. *Journal of the American Academy of Dermatology*, *89*(4), e157-e158. https://doi.org/10.1016/j.jaad.2023.05.054

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology*, *175*, 126–137. https://doi.org/10.1016/j.pneurobio.2019.01.008

Joachim von Braun, Margaret S. Archer, Gregory M. Reichberg, & Marcelo Sánchez Sorond. (2021). AI, Robotics, and Humanity: Opportunities, Risks, and Implications for Ethics and Policy. In J. von Braun et al. (Eds.), *Robotics, AI, and Humanity*. https://doi.org/10.1007/978-3-030-54173-6_1

Kumar, P., Chauhan, S., & Kumar, L. A. (2023). Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. *Engineering Applications of Artificial Intelligence*, *120*, 105894. https://doi.org/10.1016/j.engappai.2023.105894

Leeson, P. T., & Coyne, C. J. (2005). The economics of computer hacking. *JL Econ. & Pol'y*, *1*, 511.

Lesandrini, J., Idris, M. Y., & Reis, D. S. (2023). The Ethics of Artificial Intelligence and Machine Learning. *Journal of Radiology Nursing*, *42*(3), 265–266. https://doi.org/10.1016/j.jradnu.2023.05.001

Memarian, D., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, *5*, 100152, https://doi.org/10.1016/j.caeai.2023.100152

Mueller, V. C. (2012). Introduction: philosophy and theory of artificial intelligence. *Minds and Machines*, *22*(2), 67–69. https://doi.org/10.1007/s11023-012-9278-y

Palladino, N. (2023). A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy*, *47*(5), 102479. https://doi.org/10.1016/j.telpol.2022.102479

Rich, E., Knight, K., & Nair, S. (2009). *Artificial Intelligence*. Tata McGraw Hill.

Saveliev, A. M., Zhurenkov, D. A., Poikin, A. E., & Berkutova, T. A. (2021). Ethics of Artificial Intelligence and Post-non-classical Scientific Rationality. *IFAC-PapersOnLine*, *54*(13), 397–401. https://doi.org/10.1016/j.ifacol.2021.10.480

Schank, R. C. (1991). Where's the AI? *AI magazine*, *12*(4), 38. https://doi.org/10.1609/aimag.v12i4.917

Scimeca, M., & Bonfiglio, R. (2023). Comment on redefining authorship in the era of artificial intelligence: balancing ethics, transparency, and progress. *ESMO Open*, *8*(5), 101635. https://doi.org/10.1016/j.esmoop.2023.101635

Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, *16*(3), 26–33. https://doi.org/10.1109/msp.2018.2701164

Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring ethics and human rights in artificial intelligence – A Delphi study. *Technological Forecasting and Social Change*, *191*, 122502. https://doi.org/10.1016/j.techfore.2023.122502

Stahl, B. C. (2021). Perspectives on Artificial Intelligence. In *Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance* (pp. 7–17). Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9_2

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23–65). Springer Netherlands.

Xiao-Fan, L., Zhaoyang, W., Wei, Zh., Guoyu, L., & Gwo-Jen, H. (2023). Technological support to foster students' artificial intelligence ethics: An augmented reality-based contextualized dilemma discussion approach. *Computers & Education*, *201*, 104813. https://doi.org/10.1016/j.compedu.2023.104813

Zhang, Y., Wu, M., Yijun Tian, G., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, *222*, 106994. https://doi.org/10.1016/j.knosys.2021.106994

Author information

Yadav Nidhika – Ph.D., Independent Researcher
Address: Chattarpur Enclave Phase II Delhi-110074, India
E-mail: yadavnidhika68@gmail.com
ORCID ID: https://orcid.org/0000-0002-0165-6453
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57195398958
WoS ID: https://www.webofscience.com/wos/author/record/2269009
Google Scholar ID: https://scholar.google.com/citations?user=fndffSwAAAAJ

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law
PASJC: 3308 / Law
WoS: OM / Law

Article history

Date of receipt – July 23, 2023
Date of approval – October 25, 2023
Date of acceptance – November 30, 2023
Date of online placement – December 15, 2023