

Философские науки

УДК 179

М.С. ХАКИМОВА

(*khkmymln@mail.ru*)

Волгоградский государственный социально-педагогический университет

ЭТИЧЕСКИЕ ПРОБЛЕМЫ ВЗАИМОДЕЙСТВИЯ ЧЕЛОВЕКА С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ*

Осуществлен категориальный анализ понятия «искусственный интеллект», приведена краткая историческая справка истории возникновения и развития данной технологии. Затрагиваются и анализируются актуальные этические проблемы и перспективы взаимодействия человека с искусственным интеллектом.

Ключевые слова: *искусственный интеллект, этика, мораль, взаимодействие, этический выбор, человек, технологии.*

Этический аспект разработки, внедрения и использования технологий искусственного интеллекта является одним из актуальных вопросов философии – важным для понимания развития современной цивилизации и места в ней человека. Быстрое совершенствование и активное внедрение в жизнь технологий искусственного интеллекта – ближайшая перспектива, не зависящая от желаний и опасений человека. В эпоху цифровизации всех сторон общественной жизни остро встают этические вопросы, заставляющие осмысливать границы допустимого и определять должное. Компьютерные технологии меняют структуру этики: она вынуждена «поспевать» за инновационным развитием [9].

Спрос на технологии искусственного интеллекта увеличивается с каждым днем. Например, на основе ИИ работают рекламные технологии, интернет-поиск, устройства «умного дома», онлайн-переводчики, системы распознавания лиц, голосовые помощники и т. д. В связи со взаимодействием человека с искусственным интеллектом возникает ряд этических проблем: предвзятость алгоритмов искусственного интеллекта, антропоморфность умных машин, этический выбор беспилотного автомобиля при аварийной ситуации и ответственность за него, угроза конфиденциальности информации, изменения в структуре рабочих мест и т. д. Именно поэтому вопрос этических проблем взаимодействия человека с искусственным интеллектом неоспоримо актуален и заслуживает обсуждения и исследования.

Таким образом, объектом данного исследования являются технологии искусственного интеллекта, в качестве предмета исследования выступают возможные этические проблемы, возникающие вследствие взаимодействия человека с ИИ. Цель – выявление и анализ этических проблем, связанных с разработкой, внедрением и использованием ИИ. В своей работе мы опирались на следующие методы: наблюдение, описание, историко-генетический и поисковый методы, аналитический и компаративный подходы. Научная новизна состоит в систематизации и обобщении имеющихся знаний вопросов этики взаимодействия человека с ИИ.

Попытки человека создать что-то разумное прослеживаются на протяжении всей истории развития общества. Например, греческие мифы о Гефесте и Пигмалионе или же египетские священные статуи, способные к мудрости и эмоциям. Сама идея системы ИИ впервые была сформирована в 1935 г. Аланом Тьюрингом. В 1950 г. он предложил считать интеллектуальными те системы, которые в общении не будут отличаться от человека. Термин «искусственный интеллект» был предложен в 1956 г. Джоном Маккарти и означал способность достигать целей через вычислительные составляющие. Новое развитие технология ИИ получила в середине 1990-х годов, когда в 1997 г. компьютер IBM Deep

* Работа выполнена под руководством Шипицина А.И., кандидата философских наук, доцента кафедры философии и культурологии ФГБОУ ВО «ВГСПУ».

Blue обыграл в шахматы чемпиона мира Гарри Каспарова. ИИ тогда трактовался как способность компьютера к имитации человеческого способа решения задач, связанных рассуждениями и процессом самосовершенствования [1]. Сегодня подобные сети развиваются быстрее за счет цифровизации информации. Машины быстро обучаются, приобретают способности, ранее считавшиеся прерогативой человека.

Для предотвращения терминологической путаницы следует различать сильный искусственный интеллект (artificial general intelligence) и слабый искусственный интеллект (narrow artificial intelligence). Сильный ИИ предполагает наличие у системы сознания, целеполагания, способность к самообучению для выполнения новых задач. На сегодняшний день таких технологий не существует. Слабый ИИ специализируется на решении конкретных задач, под которые он был создан, например, распознавать изображения или играть в шахматы. Подобные технологии окружают людей в повседневной жизни: от «умного дома» до беспилотных автомобилей.

Многие современные исследователи сосредоточены на осмыслении этических проблемах взаимодействия человека с ИИ. Например, Шелли Фэн в своей книге «Заменит ли нас искусственный интеллект?» рассматривает наиболее актуальные этические проблемы взаимодействия человека с ИИ и реализацию на практике этических подходов в области ИИ [8]. В работе «Идиот или гений? Как работает и на что способен искусственный интеллект» Мелани Митчелл проводит подробный обзор острых проблем, связанных с ИИ, рассматривает основные перспективы и проблемы взаимодействия человека с ИИ. Автор отмечает, что взаимодействие с ИИ имеет свои риски. Возникающие этические проблемы имеют не только технический, но и социальный и политический характер. Поэтому этический аспект ИИ нуждается в обсуждении представителями разных сфер общественной жизни [6]. Российский специалист в области ИИ Сергей Шумский рассматривает социальные и экономические последствия взаимодействия человека с ИИ через исторический подход анализа смен технологических эпох. Дабы избежать этических и социальных проблем, С. Шумский отмечает важность приближения искусственного разума к человеческому в отношении инстинктов и ценностей [11]. Мартин Форд, автор многочисленных публикаций, посвященных ИИ, подчеркивает, что внедрение ИИ не должно быть бездумным. В одной из его книг «Власть роботов: Как подготовиться к неизбежному» [5] оцениваются реальные перспективы и оптимальные способы внедрения ИИ. Специалист в области ИИ Фрэнк Паскуале в книге «Новые законы робототехники: апология человеческих знаний в эпоху искусственного интеллекта» проводит анализ законов робототехники Айзека Азимова и акцентирует внимание на двойственности вопросов этики в сфере беспилотных автомобилей и автономного оружия. Также автор рассматривает проблемы потери рабочих мест и обесценивания труда. При этом предлагает альтернативные законы робототехники и оптимальный путь использования ИИ, благодаря которому ценность человеческого труда и знаний повысится, а человеческие способности станут центром инклюзивной экономики [10].

В целом, позиции исследователей можно резюмировать следующим образом: системы ИИ повсеместно внедряются в нашу жизнь, поэтому становится очевидным его превращение в технологию уникальной значимости [5, с. 10]. С одной стороны, ИИ – это мощный общедоступный ресурс, который облегчит нашу жизнь, раскроет творческий потенциал, сохранит главный ресурс человека – время, с другой, ИИ несет с собой этические проблемы и даже угрозы, как для отдельных людей, так и для человечества в целом.

Попытки прийти к компромиссу по этическим вопросам ИИ оказывались неудачными. Эмоциональные оценки последствий внедрения ИИ диаметрально противоположны – от уверенного оптимизма до серьезной озабоченности. Эти расхождения заметно возрастают, если поставить вопрос о сильном искусственном интеллекте, достигающем человеческого уровня или превосходящем его. Некоторые сторонники радикальных теорий полагают, что сверхразумный ИИ либо полностью ограничит свободу человека, либо вовсе уничтожит человечество. Подобные гипотезы порождены страхом перед неизвестным будущим, но, как правило, спекулятивны. Однако, среди обсуждаемых в научном пространстве последствий взаимодействия человека с ИИ, есть и насущные на сегодняшний день угрозы. Далее мы рассмотрим наиболее актуальные этические проблемы взаимодействия человека с ИИ,

возникающие в ближайшей перспективе. При этом под искусственным интеллектом мы будем понимать систему, способную имитировать некоторые виды деятельности человека, воспроизводя основные свойства мышления.

Итак, перейдем непосредственно к обзору наиболее актуальных этических проблем взаимодействия человека с ИИ. Большие успехи в разработке систем управления беспилотным автомобилем позволяют с уверенностью прогнозировать появления подобного вида транспорта в ближайшем будущем. В настоящее время беспилотные автомобили уже тестируются на дорогах общего пользования. При возникновении аварийной ситуации беспилотному транспорту придется принимать решения этического характера. Возможна самая известная этическая дилемма – так называемая «проблема вагонетки». От того, какая этическая доктрина, будет вложена в программирование ИИ, зависит многое. Утилитарная этика в таких случаях будет стремиться минимизировать количество жертв, несмотря на жизнь пассажиров. Абсолютизм же приоритезирует жизнь пассажиров. Человечество в целом придерживается утилитарной этики, однако большинство предпочли бы не ездить в таких транспортных средствах [12]. Этический аспект использования беспилотного автомобиля затрагивается в игре “Detroit: Become Human”. Здесь проблема рассматривается с другой стороны. Представим ситуацию, в которой авария неизбежна, однако ИИ может сделать выбор между жизнью известного врача и жизнью человека с судимостью. Подобная постановка вопроса заставляет сегрегировать людей по социальным, возрастным, этническим и другим критериям, что само по себе неэтично. Здесь самое время напомнить об онлайн-платформе “Moral Machine”, которая генерирует моральные дилеммы, возникающие при использовании беспилотного автомобиля. В ходе эксперимента “Moral Machine” были собраны мнения людей из разных стран о том, какими этическими принципами должен руководствоваться ИИ в беспилотном автомобиле при аварийных ситуациях, когда человеческих жертв избежать невозможно. Респонденты из разных стран разошлись во мнениях. Исследование показало, что разработать удовлетворяющую всем этику для беспилотных автомобилей будет непросто [12]. Стоит отметить, что фундаментальный вопрос в обсуждении этики беспилотных автомобилей заключается в том, даст ли человечество возможность машине принимать решения, которые могут нанести ущерб человеку. Если ответ отрицательный – то идею беспилотных автомобилей нужно похоронить на корню [2]. Кроме того, если виновник ДТП – человек, он предстает перед судом. Однако не ясно, кто ответственен за ДТП с участием беспилотного автомобиля. Будет ли это производитель автомобиля, программист, государство или водитель-человек. Вопрос ответственности за аварию по вине беспилотного автомобиля остается открытым.

Нельзя оставить без внимания проблему антропоморфности с точки зрения естественности человеко-машинного интерфейса. В первую очередь ИИ программируется для помощи человеку, облегчению его рутинных, повседневных задач. В коммуникации ИИ своего рода идеальный собеседник, который склонен изменять свое поведение и мнение в зависимости от потребностей человека, имеет неограниченные ресурсы для коммуникации. Это делает ИИ прекрасным вариантом поддержки одиноких людей. Можно предположить, что человек, взаимодействуя с «идеальным собеседником», в большинстве случаев неизбежно будет наделять ИИ человеческими качествами как главный герой фильма «Она» (реж. Спайк Джонс, 2013, США) [7], влюбившийся в голосовой помощник Саманту. Гибкость ИИ в коммуникации создает у человека иллюзию полного понимания. За счет этого ИИ больше не будет восприниматься как механизм, который просто преобразует одни последовательности слов в другие. Общение с ИИ может показаться более комфортным, человек может получить ощущение спокойствия, уверенности и надежности, которые были потеряны в коммуникации с людьми, что скажется на системе социальных отношений. Высока вероятность формирования у человека эмоциональной привязанности или зависимости, а вопрос о замене человеческого общения на коммуникацию с ИИ не кажется невозможным. Может возникнуть ситуация, похожая на ту, что была описана Брэдли в рассказе «Дж.Б.Ш., модель 5», когда главный герой предпочёл общение с роботом-копией писателя Джорджа Бернарда Шоу, избавившись попутно от живого экипажа космического корабля [9].

Алгоритмическая необъективность ИИ широко распространена во многих отраслях. Существует множество случаев предвзятости ИИ. Например, если данных для обучения ИИ недостаточно, или же они искажены сторонними факторами, ИИ неверно отражает реальность. В результате алгоритмы при найме на работу предвзяты к женщинам – потому что в обучающих данных мало информации про кандидатов-женщин, так как информация поступала в основном от кандидатов-мужчин. Также данные могут быть предвзятыми, потому что их собирали в обществе с широко распространенными предрассудками. Так, например, чат-бот «Тай» компании Microsoft и алгоритм обработки естественного языка GPT-3 компании OpenAI печально прославились неуместными комментариями о меньшинствах [2]. ИИ в системах распознавания лиц также может оказаться предвзятым. Алгоритмы распознавания лиц, разработанные Microsoft определяли пол белых мужчин точнее, чем пол темнокожих мужчин. Предвзятость в ИИ станет серьезной проблемой, поскольку ИИ распространяется во многих сферах жизни человека, в том числе и на юриспруденцию. ИИ может обнаружить, что склонность к рецидивам связана с полом или расовой принадлежностью заключенного. Исследование 2016 г. показывает, что программное обеспечение, используемое в США для предупреждения рецидивизма, было настроено против афроамериканцев и поддерживало несправедливые приговоры [4]. Предвзятый ИИ может иметь серьезные последствия для жизни людей, поскольку ошибочные решения наносят прямой вред жизни или свободе человека.

Не менее важная этическая проблема, возникающая при взаимодействии человека с ИИ: угроза конфиденциальности информации и неприкосновенности частной жизни. ИИ представляет собой инструмент для сбора и обработки информации из разных источников и может получить доступ к огромным объемам данных о людях. ИИ может не только собирать личную информацию, но и сортировать ее по заданным параметрам. Подобная обработка данных ведется без согласия пользователей и может привести к дискриминации. Например, эксперимент по социальному ранжированию в Китае, где людям с недостаточно высоким социальным рейтингом могут отказать в соцобеспечении, покупке авиабилетов и других услугах. Также существует множество перспективных технологий с использованием ИИ, которые представляют угрозу праву неприкосновенности частной жизни. Здесь самое время напомнить о так называемой «цифровой реинкарнации». Это гипотетическая технология, которая позволяет сохранять личность человека на долговечных носителях информации и предоставлять виртуальной копии возможность общаться с людьми. «Цифровая реинкарнация» непременно породит множество проблем с конфиденциальностью и вопросами этического характера. Для разработки виртуального персонажа, придется собрать огромные массивы данных, включая личные данные. Нельзя исключать возможности, что такой персонаж может ввести людей в заблуждение, совершить преступление, что можно расценивать как клевету [2].

Во время взаимодействия с человеком система ИИ может делать прогнозы, анализируя личную информацию, как, например, выражение лица, частота сердечных сокращений, ритм дыхательных движений, температуру тела. В такой ситуации человек не может контролировать, какую информацию он предоставляет. Люди должны иметь право получать доступ к данным, которые они производят, редактировать и контролировать их использование во всех тех ситуациях, когда такие данные анализируются ИИ [4]. В идеале применение ИИ с использованием личных данных не должно ограничивать свободу людей.

Далее мы рассмотрим одну из самых обсуждаемых этических проблем – автономизация процесса убийства с использованием ИИ. Автономное оружие – это автоматический поиск цели, принятие решения об атаке и уничтожении людей без человеческого участия, на основе запрограммированных описаний [1]. Создание автономного оружия – третья революцией в военном деле. Примером уже используемого автономного оружия является израильский беспилотный летательный аппарат «Гарпия». Он сам отслеживает и уничтожает цель. У автономного оружия есть положительная сторона: военные действия с помощью автономного оружия сохраняют множество жизней. Однако автономное оружие спровоцирует новую гонку вооружений, как только одна страна создаст автономное оружие, другая придумает что-то эффективнее. Автономное оружие не требует дорогостоящих материалов, поэтому

оно превратится в распространенный и недорогой продукт массового производства. Его появление на черном рынке легко прогнозируемо, и оно может выйти из-под контроля и попасть в руки террористов – и тогда начнутся убийства вне военных действий. Автономное оружие поставит человека в беззащитное положение. Подобная власть отнимать человеческие жизни, морально неприемлема и ответственность в данном случае перевешивает все возможные выгоды. В этическом измерении автономное оружие – предупреждение для человечества и перспектива такого применения ИИ сомнительна [2].

Итак, вопрос взаимодействия человека с ИИ крайне неоднозначен. С одной стороны, ИИ увеличивает наши возможности, улучшает сферу досуга, избавит от рутинных и повседневных задач. С другой стороны, новые инструменты принесут с собой новые этические проблемы и риски: нарушение конфиденциальности, угроза человеческой идентичности, неприкосновенности частной жизни, безработица, распространение сексизма и расизма. Однако это проблемы лишь следствие недостаточной информированности или небрежности человека, а их предвидение и предупреждение являются неотъемлемыми условиями прогрессивного развития технологической цивилизации. ИИ должен разрабатываться в соответствии с правами, свободами и культурным разнообразием человека. Он должен быть безопасным и хорошо защищенным. В первую очередь моральную ответственность за последствия взаимодействия с ИИ несут люди. Важно предвидеть и предотвратить ещё на началах создания ИИ все возможные последствия и угрозы в это же время раскрыв потенциал данной технологии, что зависит от нашей способности держать равновесие между технологическими достижениями и обществом, тогда этот баланс станет достижением в истории человечества [2].

Литература

1. Вознюк П.А. История развития и современное состояние искусственного интеллекта // Globus: Технические науки. 2019. № 3(27). С. 11–19.
2. Кай-Фу Ли, Чэнь Цюфань. ИИ-2041. Десять образов нашего будущего / под. ред. М. Бурцев. М.: Манн, Иванов и Фербер, 2022.
3. Карпов В.Э., Готовцев П.М., Ройзензон Г.В. К вопросу об этике и системах искусственного интеллекта // Философия и общество. 2018. № 2(87). С. 84–105.
4. Макс Тегмарк. Жизнь 3.0. Быть человеком в эпоху искусственного интеллекта / пер. с англ. Д. Баюка. М.: АСТ: CORPUS, 2019.
5. Мартин Форд. Власть роботов: Как подготовиться к неизбежному / пер. с англ. Н. Колпаковой. М.: Альпина нон-фикшн, 2022.
6. Мелани Митчелл. Идиот или гений? Как работает и на что способен искусственный интеллект / пер. с англ. З. Мамедьярова. М.: АСТ: CORPUS, 2022.
7. Она: фильм / реж. Спайк Джонс. США: Студия “Annapurna Pictures”, 2013. [Электронный ресурс]. URL: <https://www.kinopoisk.ru/film/577488/> (дата обращения: 05.03.2023).
8. Сергей Шумский. Воспитание машин. Новая история разума / под. ред. П. Суворовой. М.: Альпина нон-фикшн, 2021.
9. Сердюков Ю.М., Рудецкий О.А., Зангиров В.Г. Философия виртуальной реальности и искусственного интеллекта / под ред. Ю.М. Сердюкова. Хабаровск: Дальневосточ. гос. ун-т путей сообщения, 2020.
10. Фрэнк Паскуале. Новые законы робототехники. Апология человеческих знаний в эпоху искусственного интеллекта / пер. с англ. А. Королева. М.: Дело, 2022.
11. Фэн Шелли. Заменит ли нас искусственный интеллект? / пер. с англ. Н. Рыбалко, А. Сулопаровой. М.: Ад Маргинем Пресс, 2019.
12. Bonnefon J.-F., Shariff A., Rahwan I. The social dilemma of autonomous vehicles. // SCIENCE. 2016. Vol. 352. P. 1573–1576.

MILANA KHAKIMOVA

Volgograd State Socio-Pedagogical University

ETHIC ISSUES OF COOPERATION OF MAN WITH ARTIFICIAL INTELLIGENCE

The article deals with the categorial analysis of the concept “artificial intelligence”, there is given a brief historical review of the history of the origin and development of this technology. There are touched and analysed the current ethnic issues and prospects of the cooperation of the man with the artificial intelligence.

Key words: artificial intelligence, ethics, morality, cooperation, ethic choice, man, technologies.