

ЗНАЧИМОСТЬ ЧЕЛОВЕЧЕСКИХ ПРАВСТВЕННЫХ ЦЕННОСТЕЙ ДЛЯ ВОВЛЕЧЕНИЯ ПОДРОСТКОВ В ФОРМИРОВАНИЕ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАЗОВАТЕЛЬНОЙ И СОЦИАЛЬНОЙ СФЕРАХ

Донец Т.А.^{1,2}

¹ГАОУ ВО «Московский городской педагогический университет», Москва;

²ГБОУ Школа № 1228 «Лефортово», Москва, e-mail: tdonets2010@yandex.ru

Целью данной работы представляется обоснование значимости и выявление нравственных ценностей, актуальных для рассмотрения и принятия в качестве фундаментальных оснований в ходе разработки этических принципов инжиниринга и функционирования систем искусственного интеллекта образовательной сферы применительно к процессу социализации и привлечения подростков к социальной деятельности. Конкретизация значимых установленных этических ценностей, соответствующих интересам человечества, содействующих воспитательному воздействию на подростков и развитию общества, была достигнута на основе научного метода изучения и анализа научной литературы и проведения опросов и анкетирования при реализации следующих задач исследования: (1) анализе теоретических подходов и понятий нравственных ценностей в контексте технологии искусственного интеллекта, (2) систематизации человеческих нравственных ценностей, ключевых для включения в этикет искусственного интеллекта сферы образования и подростковой социализации, (3) оценке вызовов и потенциальных рисков, которые могут привести к негативным последствиям использования искусственного интеллекта в образовательных целях, (4) изучении предлагаемых на сегодняшний день практик привития нравственных ценностей и разработке рекомендаций по приобщению подростковой аудитории к нравственным ценностям, которые посредством их участия будут в будущем внедряться в этические нормы искусственного интеллекта в ходе проектирования новых и обновления предыдущих версий технологий на основе искусственного интеллекта и в ходе оценки и контроля этичности систем искусственного интеллекта на определенном этапе социального развития. Полученные результаты будут способствовать применению подходов с проявлением более высокого уровня ответственности и гуманизма, справедливости и устойчивости общественного развития. Итоговые выводы и рекомендации представляют интерес для педагогических работников и разработчиков искусственного интеллекта, осуществляющих разработки для образовательных учреждений, а также других участников образовательной деятельности, заинтересованных в интеграции разработки этических принципов использования искусственного интеллекта в молодежные образовательные практики.

Ключевые слова: этика искусственного интеллекта, общечеловеческие ценности, обучающий искусственный интеллект, нравственные принципы, совместная деятельность, образовательные практики, социальное устройство.

HUMAN MORAL VALUE IMPORTANCE FOR THE ARTIFICIAL INTELLIGENCE ETHICS ELABORATION VIA ADOLESCENT PARTICIPATION IN EDUCATIONAL AND SOCIAL SPHERES

Donets T.A.

Moscow City University, Moscow;

School No. 1228 "Lefortovo", Moscow, e-mail: tdonets2010@yandex.ru

The purpose of the research is to substantiate the significance and identify moral values that are relevant for consideration and adoption as fundamental bases in the development of ethical principles and artificial intelligence systems functioning in the educational sphere in relation to the process of socialization and involvement of adolescents in social activities. The specification of significant established ethical values that correspond to the interests of humanity, promote educational impact on adolescents and development of society was achieved on the basis of the scientific method of studying and analyzing scientific literature and conducting surveys and questionnaires in the implementation of the following research objectives: (1) analysis of theoretical approaches and concepts of moral values in the context of Artificial Intelligence technology, (2) systematization of human moral values that are key for inclusion in the etiquette of artificial intelligence in the sphere of education and adolescent socialization, (3) assessment of challenges and potential risks that can lead to negative consequences of using artificial intelligence for educational purposes, (4) study of currently proposed practices of instilling moral

values and development of recommendations for introducing adolescent audiences to moral values, which, through their participation, will in the future be introduced into the ethical standards of artificial intelligence in the course of designing new and updating previous versions of technologies based on artificial intelligence and in the course of assessment and control of the ethics of artificial intelligence systems at a certain stage of social development. The obtained results contribute to the application of approaches with a higher level of responsibility and humanism, justice and sustainability of social development. The final conclusions and recommendations are of interest to teachers and developers of artificial intelligence implementing developments for educational institutions, as well as other participants in educational activities interested in integrating the development of ethical principles for the use of artificial intelligence into youth educational practices.

Keywords: ethics of artificial intelligence, universal values, educational artificial intelligence, moral principles, joint activities, educational practices, social structure.

Введение

Глобальное внедрение технологий искусственного интеллекта (ИИ) в экономику, бизнес, политику, социальную сферу, образование поставило вопрос о формулировке этических принципов, норм и ценностей при его разработке, внедрении, использовании для предотвращения опасностей, рисков, снижения угроз для пользователей и всех вовлеченных сторон, например таких как нарушение конфиденциальности и личного пространства, частной жизни, «врачебной тайны», дискриминация личности или социальной группы, несправедливость принятого решения и многое другое. Самыми активными сначала пользователями, а в будущем – разработчиками и лицами, принимающими решения по этике ИИ, являются подростки; таким образом, они, как «ключевые фигуры» ИТ-прогресса, должны быть с самого раннего этапа вовлечены в формирование этических принципов ИИ, которые в первую очередь опираются на общечеловеческие нравственные ценности. Этическая культура ИИ в цифровую эпоху становится ключевым фактором успешного эффективного социального взаимодействия.

Именно поэтому актуален комплексный анализ особенностей внедрения и использования ИИ в российском образовательном пространстве, выявление существующих этических принципов ИИ сферы образования и недостающих их составляющих, поиск новых образовательных практик и модернизация применяемых, которые бы, с одной стороны, способствовали развитию нравственных ценностей молодежи, а с другой, в ходе совместной социальной деятельности приводили к выработке требуемых элементов этики искусственного интеллекта.

Наиболее широко искусственный интеллект в сфере образования стал применяться с начала XXI в. в связи с реализацией программы устойчивого развития в глобальном масштабе, предложенной ЮНЕСКО и одобренной правительствами большинства стран мира. В государственные и международные системы образования были внедрены инструменты на основе искусственного интеллекта на следующих основаниях включения субъектов обучения: с ориентацией на обучающихся (автоматизация навыков, поиск источников и материалов, доступ к базам данных), с ориентацией на преподавателей (проверка на плагиат,

автоматизация оценивания работ обучающихся, организация обратной связи по факту обращения), с равной ориентацией на всех бенефициаров образовательной деятельности (базы данных источников, базы данных прогресса обучения, МООКи, тренажеры, тестирование навыков и программы на их совершенствование). Кроме программ для непосредственного взаимодействия между участниками образовательного процесса на основе ИИ, искусственный интеллект был также введен в процессы организации и управления процессом обучения. Несмотря на то, что искусственный интеллект проявил себя как эффективный алгоритм автоматизации при выполнении рутинной работы, высвобождающий время преподавателей для коммуникации с обучающимися для решения задач повышенной сложности на более глубоком уровне освоения содержания образования, его внедрение во многих странах, включая Россию, вызывает споры и недовольство со стороны как педагогической, так и родительской общественности. Являясь дорогостоящей технологией, ИИ разрабатывается командами, в состав которых включены программисты и иные технические специалисты, в итоге появляются продукты, на основе возможностей разработчиков и доступа к имеющимся технологиям, но не учитывающие потребности и особенности обучающихся, «оторванные» от актуального содержания, без учета этических принципов, ценностей, юридических норм, не говоря о национальных особенностях и традициях. В некоторых странах, включая РФ, ИИ стал применяться для проверки эссе и сочинений, было заявлено, что нейросети такого направления повысят качество проверки и оценки работ обучающихся, так как они способны обнаруживать ошибки лучше, чем учителя и эксперты, тем не менее продукты, предназначенные для проверки текстов на русском и иностранных языках, пока не достигли заявленной цели, так как возникает много спорных вопросов использования стилистических приемов, особенно юмора, сатиры, а также вопросы использования лексики, которые системы считают некорректными, несмотря на правильное употребление авторами, и, наоборот, не распознают совершенных ошибок, так что после автоматизированной проверки требуется проверка «вручную», тогда как чат-боты с заявленными простыми функциями проверки односложных ответов, предоставления статистики за определенный период, оценки учебного прогресса, отбора источников для изучения с частичной автоматизацией коммуникации в приложениях и на платформах учреждений образования сократили время на подготовку к занятиям и на организацию процесса обучения как для учителей, так и для обучающихся. В некоторых странах, например в Китае, в обиход вошли смарт-камеры, отслеживающие концентрацию и внимание обучающихся по выражению лица, а также автоматизирующие контроль дисциплины во время уроков и оценивающие эффективность работы учителя.

Таким образом, можно констатировать, что в настоящее время не удалось избежать нежелательной предвзятости (человеческого фактора) при оценивании учителем, который из-

за личных предпочтений может снижать или завышать оценки, так как автоматизированное оценивание искусственным интеллектом так же несовершенно и требует повторной проверки человеком, именно поэтому продуктивная часть всех экзаменов и испытаний до сих пор проверяется вручную. Тем не менее использование программного обеспечения при анализе данных результатов тестов и других работ обучающихся позволяет выявить наиболее проблемные для усвоения темы и навыки, требующие дополнительной отработки. Другими образовательными ИИ-продуктами являются: библиотеки с обучающими материалами и тренажерами с круглосуточным доступом, индивидуальные программы под потребности конкретного обучающегося (адаптивные персонализированные треки), курсы, предлагающие по выбору синхронное или асинхронное обучение на образовательной онлайн-платформе с целью устранения пробелов и оптимизации обучения, приложения-репетиторы [1].

Проанализировав применение данных продуктов в сфере образования, можно выделить следующие преимущества: 1) улучшенная организация и структурирование материала и баз данных, 2) ускоренная автоматическая категоризация, поиск и классификация ресурсов, 3) возможность сократить время на создание учебных материалов, пособий, разработок, планов, программ, 4) образовательная индивидуализация с учетом уровня подготовки, интересов и потребностей, 5) улучшенные возможности обучения людей с ограничениями по здоровью (например, распознавание устной или письменной речи для людей с трудностями в письме и артикуляции) – и такие недостатки, как: 1) высокая стоимость программ ИИ, оборудования, обслуживания, обновления, 2) высокая энергозатратность, 3) сложный и длительный процесс тестирования, внедрения и адаптации, 4) обеспечение требуемого уровня конфиденциальности, 5) высокая вероятность несанкционированного доступа к личным данным обучающихся и учителей, 6) использование личных данных в корыстных целях, 7) сокращение личного и личностного общения, 8) нарушение связей, формирующих модели социального поведения, 9) невозможность работы обучающего ИИ со «сложными обучающимися» (алгоритмы ИИ не учитывают все личностные и психофизические особенности в комплексе), 10) невозможность просчета программой рисков и ограничений для каждого образовательного учреждения, 11) неспособность обеспечить полную инклюзивность, равенство, одинаковое качество систем данных искусственного интеллекта, 12) отсутствие универсальных этических принципов, регулирующих сбор, передачу, распространение данных образовательных систем [2].

Именно последний из недостатков – самый многоаспектный вызов образовательному научному сообществу сегодняшнего времени помимо необходимости подготовки программ искусственного интеллекта к «пониманию и осознанию» этапов и задач образовательного процесса со всей сложностью, непрерывностью и потребностью в постоянном

реформировании. Сегодняшний этап использования ИИ в образовании является вспомогательным, в перспективе заявлен основной характер использования образовательного искусственного интеллекта, который сможет обеспечить прорыв секторов «обучение – преподавание», «обратная связь – персонализированная поддержка» [3].

Так как к настоящему моменту не предложено решений по применению искусственного интеллекта, которые смогли бы вывести образование на качественно новый уровень, возникла необходимость отбора практик, которые, с одной стороны, усовершенствовали бы навыки управления своими знаниями и образованием, решения текущих проблем, критической оценки и мышления, творческого переосмысления подростков, а с другой стороны, обеспечили бы вовлечение и участие подростков в разработке этики ИИ, в нахождении баланса между преимуществами и недостатками искусственного интеллекта так, чтобы гарантировать общую пользу для системы образования как целого и для каждого обучающегося в отдельности при его использовании в качестве инструмента повышения качества образования.

Современные исследователи изучили различные аспекты, которые необходимо учесть при отборе образовательных практик участия подростков в процессе выработки этики современного ИИ.

С. Беннет исследовала этические и социальные последствия взаимодействия с реальными приложениями ИИ, сопоставляя их с положениями руководств для новых приложений методом опроса пользователей и специалистов, оценивших итоги своих разработок, как активных моральных агентов, определив степень поддержания значимого применения заложенных при проектировании и разработке этических принципов ИИ [4]. А. Оуи и С. Баум предложили помимо общечеловеческих ценностей и этических принципов в области этики искусственного интеллекта также учитывать ценности и интересы нечеловеческих существ (другие биологические виды животного и растительного мира и сам ИИ), по их мнению, нечеловеческие существа заслуживают уважения, им следует оказывать достаточное внимание, перестать игнорировать их самоценность и заняться исследованием их морального статуса и учетом непреднамеренных эффектов, оказываемых на них системами ИИ [5]. М. Ашок с коллегами проанализировал социально-технические преобразования, произошедшие в ходе использования ИИ и цифровых технологий (ЦТ) из-за недостаточно глубокой проработки этики ИИ, ограничившись лишь ключевыми принципами. Данному коллективу ученых удалось выявить этические принципы ИИ высокого уровня, 14 типов последствий применения цифровой этики, архетипы ИИ и предложить новую онтологическую структуру ИИ и виды опосредованного влияния на общество и общественную жизнь государства [6]. Ван де Поль в своей работе предложил способ определения, понимания

конкретных этических принципов и ценностей (включая уважение принципа автономии, недопустимость использования вредоносного ПО, справедливость принятых решений, объяснимость, прозрачность, подотчетность) социотехническими системами искусственного интеллекта и способ подтверждения воплощения этих ценностей в виде проектной деятельности, результат которой направлен на внедрение универсальных ценностей в системы ИИ, элементами которых являются технические артефакты, человеческие агенты, организации и институты, искусственные агенты и технические нормы регулирования взаимодействия между искусственными агентами и другими элементами. В работе автором были рассмотрены конкретные проблемы и возможности эффективного внедрения ценностей в алгоритмы ИИ [7].

П. Шренк, В. Мюллер и Л. Кайзер – еще одна группа исследователей, которых заинтересовал вопрос нравственности искусственного интеллекта, степень его приемлемости в зависимости от способности выполнять задачи в соответствии с моральными нормами и приписываемым ИИ социальным статусом, моральная приемлемость агентов ИИ (люди или компьютерные программы). Было выявлено экспериментально, что ИИ образовательных организаций с высоким статусом считается более приемлемым с моральной точки зрения, чем ИИ образовательных организаций с низким статусом, при использовании одинаковых программ, таким образом, организации образования с высоким социальным статусом получают моральные преимущества, и их высокое структурное положение в культурной иерархии усиливается вне зависимости от истинной эффективности [8].

Р. Хастути в своем исследовании всесторонне проанализировал этические аспекты эпохи искусственного интеллекта на основе библиометрии и VOSviewer-визуализации, систематизировал научные источники по преобладанию этических аспектов ИИ в образовании и по новым тенденциям этики ИИ, выявил возможности сетевого сотрудничества в рамках междисциплинарного подхода решения этических проблем в соответствии с человеческими ценностями в ходе динамичной эволюции ИИ и его этики, уделяя внимание различным аспектам: от влияния на общество до управления заинтересованными сторонами [9].

Для того чтобы технологии и системы искусственного интеллекта смогли принести человечеству наибольшую пользу и комфорт с наименьшим количеством проблем, прежде всего этического плана, необходимо, чтобы в процесс формирования этики искусственного интеллекта было привлечено как можно больше представителей разных общественных групп и возрастов, так как формирование этики будет происходить непрерывно в ходе делегирования системам искусственного интеллекта все больших полномочий и решения все более сложных задач в разных сферах человеческой жизнедеятельности; соответственно,

требуется более детально остановиться на значимости нравственных ценностей для этики искусственного интеллекта.

Цель исследования – обоснование значимости моральных принципов и нравственных ценностей и выявление актуальных нравственных ценностей в процессе привлечения обучающихся подросткового возраста к разработке этики искусственного интеллекта в сферах образования и социального развития общества.

В ходе исследования подтвердились следующие выдвинутые гипотезы: 1) сформированность нравственных ценностей подростков способствует более осознанному и ответственному использованию ИИ в образовательной и иных сферах их жизни; 2) подростковая социальная совместная деятельность по работе с ИИ и выработке новых положений этики искусственного интеллекта активизирует их участие и роль в социальном переустройстве и позволяет смоделировать новые практики этического взаимодействия; 3) осознанное применение ИИ подростками повысит учет культурных, возрастных, образовательных особенностей, уровень цифровой грамотности, критического мышления, социальную ответственность, став инструментом обновления этических принципов ИИ по мере необходимости.

Материалы и методы исследования

Исследование в рамках данной публикации опиралось на научные методы изучения и сравнительного анализа нормативных документов, экспертных интервью и научной литературы, проведения опросов и анкетирования в рамках комплексного и систематического подходов.

Анкетирование было проведено в двух общеобразовательных школах г. Москвы в 2021–2024 гг. по параллелям анонимно среди подростков 11–17 лет, опрос проводился силами обучающихся (обучающиеся одного класса параллели опрашивали обучающихся других классов). Целью анкетирования и опроса было выявление способов их пользовательской адаптации к коммуникации с ИИ, основных преимуществ и барьеров использования ИИ в образовательных и бытовых целях у подростков, а также аспектов, которые, по их мнению, требуют доработки с технической и этической стороны и готовности общества к повсеместной интеграции ИИ. Полученные показатели были округлены до целых чисел.

Результаты исследования и их обсуждение

В процессе опроса и анкетирования выяснилось, что 100 % опрошенных подростков 11 лет знакомы с ИИ, слышали о чатах GPT, приложениях для создания изображений (были названы Canva, Midjourney, DALL-E 4, Leonardo.AI) и хоть один раз общались с голосовыми помощниками (самый распространенный ответ: Алиса и Siri). Почти все подростки, 98 %, использовали голосовых помощников в учебных целях с запросом решить задачу, найти ответ,

перевести слово или текст с иностранного или на иностранный язык; 57 % респондентов данной возрастной группы понимают необходимость проверки информации в более надежных источниках и относятся к таким приложениям как к развлечению, не считая их «серьезным учебным инструментом». Все респонденты отметили, что на более чем половину заданных голосовым помощникам вопросов они либо не получили удовлетворяющего их ответа, либо представленные сведения не соответствовали действительности, были «однобокими», неполными или ввели их в заблуждение (на практике выяснилось противоположное предоставленной ИИ информации). Подростки с 12 лет включаются в активное взаимодействие с ИИ, ежедневно используют его при выполнении учебных и домашних заданий, для удовлетворения любопытства по предметным и бытовым областям, перевода на иностранный язык, для генерации изображений и видео, текстов, в качестве тренажеров, для развлечения. «Продвинутость» и частота использования ИИ в группе 12–17 лет от возраста не зависит, опрос показал, что интенсивность использования зависит от личных предпочтений, на которые повлияло количество гаджетов на каждого члена семьи в данной семье, степень ограничения времени пользования гаджетами, пример родителей, старших сестер и братьев и доступ к подобным приложениям. Все опрошенные подростки 12–17 лет подтвердили, что стопроцентного доверия к предоставляемой ИИ информации у них нет, 77 % не знают, где найти надежные источники для сверки или перепроверки сведений, 52 % слышали о том, что при генерации текстов и изображений ИИ использует базу данных и сгенерированный продукт является плагиатом данных и документов, составляющих базу данных ИИ. Ни один из опрошенных не обращался к ИИ для личностного близкого общения в трудных жизненных ситуациях, только в качестве шутки и развлечения, 87 % опрошенных не верят, что в будущем ИИ полностью сможет заменить человека, семью, друзей, традиционную медицину, образование. Все опрошенные, 100 %, столкнулись с мошенничеством, беспокойством о слишком большой зависимости от ИИ, нарушением конфиденциальности или утечкой данных (после которых следовали телефонные или онлайн-атаки мошенников) и считают, что его нужно дорабатывать для достижения большей надежности, пользы, решения глобальных проблем; самую большую опасность опрошенные видят в контроле ИИ над личными данными пользователей и предвзятости при принятии решений, 87 % считают, что ИИ нельзя доверять сферы охраны здоровья и медицины, финансов, банков и коммерции, бизнеса, государственного управления, творчества, инженерии и учебного мониторинга. Готовы доверить ИИ выполнение рутинной бытовой хозяйственной деятельности 90 % подростков. Сильное раздражение у опрошенных вызвали ситуации, когда правильные ответы были засчитаны как неверные при прохождении учебного тестирования с автоматизированной проверкой, приемлемые решения не были засчитаны как правильные, когда голосовые

ассистенты генерировали оценочные суждения личности подростков, в их комментариях содержались неэтичные формулировки, унижение, попытки принуждения; проявление нетерпимости, когда многочисленные разнообразные повторные запросы или промпты не приводили к более развернутым ответам или более точным изображениям или графикам, а сгенерированный ИИ продукт или оставался без изменений, или изменялся в худшую сторону (необоснованно большая потеря времени и усилий пользователями). На вопрос «Улучшило ли использование ИИ качество твою учебу?» 91 % ответили «нет», 66 % подростков ощущают, что их учебные результаты ухудшились, и осознают, что для освоения учебного материала нужны усилия и время на проработку материала, которых не хватает, поэтому продолжают пользоваться ИИ и другими ресурсами с готовыми ответами и решениями. Самыми популярными преимуществами ИИ были названы: 1) экономия времени на поиск информации (например, перевод незнакомых слов и текстов); 2) скорость выполнения заданий (задач) и подробное описание хода решений; 3) возможность изучать иностранный язык и другие предметы (помощь при восполнении пробелов изученного материала в случае пропусков, болезни), получить и отработать новый навык; 4) знакомство с другими точками зрения, идеями, рекомендациями. Все респонденты (100 %) данной опросной группы понимают, что 1) в важных вопросах доверять ИИ без проверки в других источниках нельзя; 2) ИИ, с одной стороны, дополняет общение, а с другой – сокращает живое человеческое общение, и считают, что 3) программы ИИ должны предусмотреть опцию объяснения логики принятия определенных решений; 4) в случаях предвзятости и (или) дискриминации должна быть возможность обжаловать или отменить решение, пожаловаться и перепрограммировать систему. Не считают обязательным соблюдение авторских прав 82 % опрошенных, они хотят, чтобы ИИ оставался инструментом «свободного и бесплатного» пользования, и высказываются против ограничений или регулирования со стороны государства.

Выявленные факторы, вызывающие беспокойство и опасение со стороны подростков при применении технологий искусственного интеллекта, являются предметом изучения исследователей во всем мире. В. Холмс [10] и Д. Боренштейн [11] рассмотрели острые проблемы и конфликтные ситуации, связанные с применением ИИ как жизненно важной технологии, нацеленной в сфере образования на получение высококачественных результатов с большим количеством преимуществ, но вызывающие озабоченность с точки зрения нарушений прав человека (крупномасштабный сбор и анализ персональных данных обучающихся, негативное психическое воздействие некоторых алгоритмов при оценивании, недопонимании смысла высказывания обучающихся искусственным интеллектом). Т. Дарадоумис [12], Е. Казим [13] и А. Нигама [14] разработали целостный набор этических принципов для применения в образовательном контексте, взяв за отправную точку несколько

или одну из многочисленных рекомендаций различных государственных структур, проведя систематический анализ задокументированных общих принципов ИИ, выявив надежность этих принципов и развивая дальнейшие направления формирования этики образовательного искусственного интеллекта.

Так как в работах исследователей представлены общие очертания перспективных направлений, они привлекают внимание специалистов смежных областей, стимулируют дальнейшие научные дискуссии, повышают интерес к данной тематике, могут являться основой для формирования общих практик и политики сферы ИИ образования, которые, в свою очередь, будут постоянно развиваться и продвигаться. В данный момент самым актуальным вопросом остается разработка точного механизма применения этических принципов «искусственного образования» для его доработки и улучшения, так как не достигнут универсальный консенсус по выбору наилучшей этической теории; пока основное внимание уделяется именно набору этических стандартов образовательного сектора. Основная трудность состоит в том, что основа этических принципов – абстрактные суждения человека и субъективные интерпретации, вовлечение большего числа заинтересованных сторон в образовательный дискурс создает дополнительные препоны широкому применению этических принципов и формальным, и дедуктивным способами. Тем не менее на специальных площадках, сайтах, конференциях продолжается сбор мнений, отзывов, предложений по улучшению использования ИИ в образовательных целях. Ряд исследователей (включая М. Райсса [15]) отметили, что наибольшую дискуссионность вызывает развитие неврологии уязвимых групп (маленькие дети, младшие школьники, инвалиды), связанной с использованием ИИ, в связи с чем, помимо аспекта инклюзивности, необходимо проработать критерии, смягчающие нарушения «алгоритмической несправедливости», системы управления образованием и политики в области ИИ образовательных процессов, так как они взаимосвязаны с развитием других значимых отраслей народного хозяйства, необходимо обеспечить их долгосрочное развитие.

Эффективное использование искусственного интеллекта, данных и аналитики, машинного обучения может позволить преподавателям сделать обучение более увлекательным за счет применения технологий погружения в виртуальную среду, автоматизации рутинных проектов, но для начала необходимо урегулировать и усилить «сотрудничество» человека с искусственным интеллектом, так как «прокси машины» должны будут выступать в качестве моральных посредников по очень чувствительным моральным вопросам. «Моральные имплицитные агенты» способны кодировать моральные ценности, выносить «моральные решения», алгоритм некорректных рекомендаций, неправильное распознавание или решение моральной дилеммы будут иметь моральные последствия,

особенно для уязвимых групп, с точки зрения моральной психологии, поэтому важно определить количество допустимых ошибок и меру ответственности за такие ошибки [16]. Неявным моральным машинам предъявляются экстремальные требования производительности, ожидая существенного увеличения производительности человека по сравнению с базовым уровнем, также широко обсуждается опасность того, что машины будут учиться, усиливать и узаконивать предубеждения, заложенные в человеческих решениях, на которых они обучаются, что машины унаследуют от людей предубеждения и сложность их исправления. Другими причинами, затрудняющими практическую реализацию компромиссов в руководящих принципах этической политики, является невозможность количественной оценки «широких этических принципов», таких как достоинство, неприкосновенность частной жизни и др., распределение дефицитного ресурса с пагубными последствиями для людей, которые не имеют приоритета при принятии решения о распределении. Многие конструктивные решения, эксперименты и интерпретация их результатов, стилизованные и реалистичные сценарии и их моделирование помогают измерению моральных предпочтений, так как «моральным машинам» ИИ требуется уравнивать большое количество конфликтующих ценностей или приоритетов, что снимает психологическое бремя с людей и изменяет социальные ожидания относительно принимаемого приемлемого решения.

Перспективный подход к вопросу делегирования сложных моральных решений машинам заключается в совместном рассмотрении степени консенсуса или раскола среди экспертов и «неспециалистов». Некоторые моральные вопросы, касающиеся каждого, такие как образование, требуют проведения глобальных расширенных экспериментов, на сегодняшний день используется технология быстрого распространения – «вирусности», что позволяет подключать миллионы участников и собирать их данные для всестороннего описания моральных ценностей, которые люди могут захотеть видеть встроенными в машины. Помимо общественного консенсуса необходимо применять разнообразие методов и конструкций, так как они отражают разумное моральное разногласие или предубеждения и предрассудки большинства; в таких случаях исчерпывающий набор методов и средств контроля позволит убедиться в устойчивости общественного консенсуса по всем экспериментальным схемам и демографическим показателям. В ближайшем будущем у людей согласно прогнозам будет увеличено взаимодействие с машинным интеллектом как моральными пациентами (например, с беспилотными объектами, такими как беспилотные автомобили или чат-боты) в процессе кооперативного или некооперативного сотрудничества с целью получения взаимной выгоды. Результаты исследований взаимодействия человека и машины в стимулирующих играх продемонстрировали конвергенцию, определенную степень просоциальных предпочтений в отношении машин, но сотрудничество с искусственным

интеллектом не достигло уровня сотрудничества между людьми. В результате возникает потребность в дальнейших исследованиях, направленных на улучшение сотрудничества человека и машины, не прибегая к гуманизации машин. Возникает необходимость применить новые социальные нормы, которые будут развиваться вокруг новых участников социального мира разумных машин, ИИ позволяет машинам принимать автономные решения от имени заинтересованных сторон, что повышает вероятность делегирования неэтичного поведения, чтобы дистанцировать человека от действия; ИИ предлагает посредничество в человеческом общении морально значимым образом.

Делегирование все большего количества задач агентам ИИ создает множество возможностей для делегирования неэтичного поведения машинам, так как ИИ может быть использован людьми со злыми намерениями расширения преступного неэтичного поведения, например «глубокое обучение» генеративных сетей упростило создание поддельного контента, при этом ИИ может действовать независимо и причинить вред в больших масштабах. ИИ-дипфейки могут создавать поддельные удостоверения личности, что позволяет сделать фишинговые атаки более персонализированными и разрушительными. Часто непостижимая работа алгоритмов создает двусмысленность: если позволить таким алгоритмам черного ящика выполнять задачи от своего имени, это повышает вероятность правдоподобного отрицания и запутывает механизм возложения ответственности за причиненный вред. Двусмысленность – еще один механизм, с помощью которого неэтичное поведение может быть делегировано машинам. Делегирование полномочий ИИ также может привести к моральным нарушениям без какого-либо злого умысла, делегирование морально желательных действий, таких как благотворительные пожертвования, агенту ИИ может действовать как инструмент обязательства, увеличивающий масштаб и частоту таких действий; существуют возможности делегировать агентам ИИ консультативную роль с динамическим внушением морального поведения человеку, что крайне важно для профилактики асоциального поведения подростков.

Технология ИИ «машинный маскарад» – опосредованная коммуникация, используется для изменения того, как участники образовательного процесса пишут, говорят или смотрят, чтобы изменить поведение своих партнеров, что тоже постепенно находит применение в обучении и корректирующей педагогике. В сфере образования также быстро распространяется технология машинных ответов на текстовые сообщения, электронные письма, с фильтрами изображений, улучшающие внешний вид объекта, и другими мощными и гибкими формами трансформации. Например, студенты могут использовать ChatGPT, чтобы звучать более компетентно, владельцы образовательного бизнеса – чтобы казаться более надежными, а пользователи образовательных социальных сетей – в соответствии с имиджем,

который они хотели бы проецировать, или моральными достоинствами, которые они хотели бы продемонстрировать. Еще только предстоит понять, какие материальные и репутационные выгоды могут получить люди, если использование ими машин не будет обнаружено, и какая часть этих выгод сохраняется в зависимости от того, как раскрывается информация об использовании машин. Возможности «машинного маскарада» – совершенные технологии изменения изображений, видео и голосов, которые могут изменить результаты моральных взаимодействий или поднять новые этические вопросы, такие как конфликт между уменьшением дискриминации и угрозой инклюзии; «машинный маскарад» предлагает обширную новую область исследований для моральной психологии, направленную на понимание того, как люди будут использовать технологии для изменения своего представления, либо с целью изменения результатов моральных взаимодействий, либо с целью управления своей моральной репутацией. Необходимо найти решение, как эти процессы могут быть модерированы различными формами раскрытия информации и как общество будет справляться с новыми этическими дилеммами. Вопросы восприятия плагиата ИИ, снижение или повышение доверия между группами людей в присутствии агентов ИИ, изменения дружеских человеческих отношений чат-ботами являются предметом активных исследовательских программ, еще незадаанные вопросы будут появляться по мере развития новых опций ИИ. Быстрое появление технологий, основанных на искусственном интеллекте, вступает в противоречие с моральными интуициями, сформированными культурой и эволюцией на протяжении тысячелетий.

Обоснование предположений о возможностях ИИ с глубоким пониманием структуры человеческой моральной психологии поможет подготовиться к миру, который будет разделен машинами и осложнен ими, но необходимо найти максимальное количество преимуществ, чтобы своевременно обратить их в пользу человека и его среды.

В области образования высока потребность в четких, интерактивных симуляциях и визуализациях, которые могут помочь людям выбрать свой собственный алгоритм и получить непосредственный опыт того, как неявные моральные машины могут повлиять на них.

Для эффективного применения во всех возможных структурах образования искусственного интеллекта, а также для привлечения подростков для его эффективного использования, разработки практик обучения работы с ним и участия в разработке этических норм обучающих машин необходимо учесть следующие факторы, связанные с социумом, социальным статусом и социальной деятельностью при применении технологий ИИ:

– моральное признание искусственного интеллекта со стороны человека зависит от социального статуса, типа агента и организации, на площадке которой он внедрен: чем выше

статус организации, тем более приемлем ее ИИ с моральной точки зрения, эффективность работы организации не имеет значения;

- использование искусственного интеллекта на сегодняшний день оценивается обществом как менее морально приемлемое (менее нравственное) по сравнению с использованием агентов-людей или простой компьютерной программы [8];
- люди-агенты в организациях с высоким статусом получают моральную выгоду на основе своего структурного положения в культурной статусной иерархии;
- тип агента, пол и рейтинг в организации влияют на моральные суждения, независимо от фактической способности агента выполнить задачу;
- социальный контекст важен для моральных суждений об ИИ;
- социальный статус определяется принадлежностью к культурным категориям, не зависит от намерений и мотивации субъекта, применим к людям, организациям и материальным артефактам;
- статусные убеждения определяют атрибуты для классификации объектов;
- компетентность агентов и экспертов включает инструментальные и моральные аспекты;
- статусные убеждения могут приводить к предвзятости подтверждения и служить сигналом для ненаблюдаемых качеств;
- ИИ может обладать социальным статусом как ранжированная культурная категория;
- ИИ может восприниматься как социальный субъект, что влияет на моральное признание;
- ИИ ассоциируется с технологическими мифами и нарративами с приписываемой высокой компетентностью, высокая доля населения мира сомневается в его надежности и моральных качествах, испытывает страх и недоверие к ИИ, препятствует его внедрению;
- неприятие ИИ, дающего советы или принимающего моральные решения; сходство с человеком повышает доверие к ИИ, люди предпочитают советы от ИИ только в определенных ситуациях;
- высокий статус организации снижает неопределенность и повышает доверие к ИИ, статус влияет на моральное принятие одинаково для всех типов агентов: ИИ, простых компьютерных программ и людей, статус может быть более значимым на практике, так как влияет на всех членов организации;
- ИИ и простые компьютерные программы не различаются по статусу; производительность агента, включая ошибки и прозрачность, важна для морального принятия [16, 17].

Таким образом, образовательные структуры и организации вкладывают огромные средства в изучение, внедрение и развитие искусственного интеллекта, который обладает огромным потенциалом в образовании, социализации и организации совместной деятельности подростков, а также в улучшении благосостояния отдельных подростков и подростковых объединений. Реализация этого воспитательного потенциала требует лучшего понимания различных этических вопросов по обучению, использованию, обслуживанию ИИ, делегированию ему принятия решений, имеющих моральные воздействия и последствия, притом что определение степени этичности вызывает споры (этические дилеммы, расхождения во мнениях) о правильности курса действий, что требует в настоящий момент и будет требовать в будущем подключения широких слоев населения и экспертов для вынесения решений об этичности, соответствии моральным ценностям и принципам. Вопросы этики искусственного интеллекта придется решать как отдельным индивидам, государственным учреждениям, так и общественным организациям и международным объединениям.

Для того чтобы цивилизованно решать такие проблемы, адекватно оценивать процедуру обработки данных систем ИИ, рассматривать множество этических подходов и приходиться к общественному консенсусу, максимально учитывая фундаментальные разногласия людей относительно этичного поведения, необходимо начиная с младшего подросткового возраста знакомить подрастающее поколение с теорией моральных оснований, прививать общечеловеческие моральные ценности, устои и основы, разбирать кейсы, сначала касающиеся их деятельности и их окружения, а затем переходить к обсуждению и привлекать к участию в опросах, конференциях и других форматах для того, чтобы при решении «нерешенных проблем» этики ИИ были учтены индивидуальные моральные различия на нескольких уровнях ИИ, так как помимо базовых традиционных ценностей в среднем каждые 50 лет происходит переоценка «небазовых» (текущих) моральных ценностей каждого этапа развития цивилизации [18].

Для отбора конкретных образовательных практик и форм работы, направленных на разработку принципов этики ИИ, необходимо предоставить возможность подросткам находить решения социальных проблем, которые они считают первостепенными и важными. Вторым организационным условием успешных проектов и практик социальных изменений является наличие междисциплинарного и инновационного характера деятельности. Социальное инновационное образование с учетом современных реалий и потребностей (к которым на сегодня относится формирование этики ИИ) может иметь формы практико-ориентированных программ и состоять из практик использования креативности для разработки решений по улучшению критериев этики ИИ, в конечном счете повышающих благосостояние людей и общества, развивающих компетенции для выявления возможностей

создания новых социальных ценностей, формирования сотрудничества, построения социальных отношений, принятия инновационных мер для более устойчивого развития общества. Такие учебные программы учебных заведений или организаций дополнительного образования (семестровые, каникулярные, в виде серии семинаров или как часть отдельных дисциплин), рассматривающие понятия «гражданское образование», «реформаторство», «реформатор», «социальное предпринимательство», «применение искусственного интеллекта в социальных проектах и процессах», «разработка этических норм искусственного интеллекта социальной сферы» позволят подросткам стать проводниками перемен, оказать влияние на свои сообщества своими инновационными социальными проектами, обучаться и обучать своих сверстников процессу социальных инноваций.

Практикой, не требующей основательной подготовки, является проведение опросов, интервьюирование и анкетирование; практики эмпирическо-трансформационного обучения, проведения систематических или краткосрочных исследований более ресурсо- и времязатратны, их результаты смогут стать обширной базой данных по обучению служению, гражданскому образованию, предпринимательскому образованию, обучению инициаторов перемен, внедрению новых ценностей искусственного интеллекта в системы социального сектора. Практиками продвижения социальных перемен и гражданской активности могут стать: просветительство, подвижничество, амбассадорство, тренинг, коучинг, менторинг. Когда гражданственность станет глобальным брендом усилиями различных организаций и государственных уровней, уместен переход от участия в рамках образовательных организаций к сообществам и клубам, не связанным со школами, деятельность которых отражают такие направленности, как «Новое поколение граждан», «Молодежный институт гражданственности», «Институт гражданского лидерства», «Служба общественных действий», «Работаем и внедряем вместе», «Школа и общество», «Социальные контакты подростков». Помимо таких навыков нового века, как коммуникация, сотрудничество, креативность, критическое мышление, социально активным подросткам для свершения социальных изменений потребуется развитие гражданской активности и ответственности, осведомленность о социальных проблемах и способах их решения, приверженность к вовлечению и участию в совместной деятельности по изменению социальной жизни, расширению прав и возможностей, приверженность к влиянию на общество, эмпатия, развитый эмоциональный интеллект, а для эффективного совместного участия в социальной жизни – самодисциплина, уважение общественных ценностей, гибкость, лидерские навыки, навыки презентации и самопрезентации, управления временем и ресурсами, настроенность на преодоление трудностей, эффективные достижения и совершенствование. Социализация подростков начального этапа обучения предполагает зачатки формирования

вышеперечисленных социальных навыков, для их дальнейшего развития подойдут такие практики, как проекты социальных изменений (проект по обучению служению людям, проект общественной работы, проект организации позитивных групп сверстников), могут применяться и такие формы организации, как коммуны, деловые клубы, клубы юных инноваторов этики ИИ, регулярные постоянные профильные смены в лагерях, молодежные движения со своими блогами (влогами). Подростковый активизм позволяет преодолеть трудности, добиться удовлетворенности жизни и непрерывного ее совершенствования, расширения возможностей и мотивации к дальнейшим изменениям, обрести уверенность и добиться уважения окружающих и самоуважения, а развитая социальная и этическая ответственность молодого поколения – залог изменения к лучшему, комфортного существования для всех.

Для продвижения идей подросткового социального активизма педагоги могут включать его аспекты обучения в другие предметные области, например на уроках изобразительного искусства можно научиться рисовать плакаты и рекламные материалы, на уроках математики – планировать бюджет и изучать методы исследования, на уроках экономической географии – разрабатывать маркетинговые приемы, на уроках истории – изучать причины социальных проблем, а далее, в зависимости от потребностей и способностей обучающихся, добавить в программу внеурочную деятельность, адаптировать уровень поддержки, обеспечить взаимодействие с другими сообществами и через сетевое взаимодействие организовать «большую структуру» социального подросткового новаторства, создавая возможность долгосрочного развития данного направления. Еще одним необходимым условием для успешной непрерывной социально полезной деятельности является присутствие в социальных сетях в виде канала или хаба, а для привлечения особого внимания к какому-то аспекту или проблеме используется технология «вирусного распространения»: что-то передается («репостится») от пользователя к пользователю с большой скоростью, так что сразу огромное количество пользователей сети узнает о каком-либо явлении, призыве или действии.

Вопрос вовлечения подростков в процессы социального развития и улучшения, включая развитие этики ИИ, требует постоянного пересмотра стратегий и практик из-за постоянно меняющихся обстоятельств и вводных данных, а подготовка подрастающего поколения к решению глобальных социальных проблем, которые не просто ждут их во взрослой жизни, а с которыми они часто сталкиваются уже на этапе получения образования в школе, не должна прекращаться ни на секунду.

Заключение

В исследовании были выявлены значимые актуальные нравственные человеческие и общественные ценности, играющие главную роль в реализации социального потенциала общества и интеграции в проектирование и внедрение технологий искусственного интеллекта, применяемые в образовательном процессе, и актуализированы практики приобщения подростков, являющихся агентами и субъектами процесса образования, самообразования и собственного социального обустройства. Подростки в качестве будущих субъектов социальных сообществ и структур должны быть способны анализировать постоянно меняющиеся концептуальные основы, определяя необходимость выявления и учета новых нравственных ценностей, для чего необходимо приложить усилия по воспитанию «нравственного личностного стержня» современного поколения, показав, как соблюдение этических норм на основе ценностей влияет на принятие решений и как это отражается на качестве образования, образовательных результатах, удовлетворенности пользователей и социальном взаимодействии. Практическая значимость исследования представлена рекомендациями по приобщению подростков к общечеловеческим ценностям и приобщению подростков к интеграции этих ценностей при проектировании этики ИИ, а значит, их включению в социально значимые программы и проекты. Теоретическая значимость работы заключается в выявлении социально-культурных аспектов, влияющих на приобщение и восприятие нравственных ценностей подростками, и в выявлении их роли в процессе формирования новой этики ИИ по мере дальнейшего развития этой технологии применительно к образованию, которые с учетом новых образовательных стратегий могут быть заложены в разработку обновленной концепции этики искусственного интеллекта.

Таким образом, поставленные цели и задачи данного исследования были осуществлены, выдвинутые гипотезы подтверждены, в итоге на основании вышеизложенного можно утверждать, что формирование нравственных ценностей подростков, их заинтересованность и вовлеченность в процессы реформирования этики ИИ в соответствии с изменениями является крайне важным условием устойчивого развития российского цифрового общества, превращая его в передовое общество, а совместный процесс разработки способствует повышению ответственности, социальной вовлеченности, совершенствует критическое мышление. Предложенная модель, выявленные условия эффективной совместной деятельности подростков по формированию новой этики ИИ в образовательной сфере и рекомендации по выбору практик позволят интегрировать образовательные молодежные инициативы в этико-центрированные проекты, повысить доверие к ИИ, решить вопросы справедливости, приватности и ответственности при работе с ИИ в сфере образования. В дальнейших исследованиях по данной проблематике можно сосредоточиться на выявлении этнокультурных различий восприятия этики искусственного

интеллекта, влиянии совместных проектов по разработке этических принципов ИИ на представителей разных этнических групп и формирование новых ценностей в рамках этносов, на создании образовательных программ для подростков по этике ИИ.

Список литературы

1. Tolmeijer S., Kneer M., Sarasua C., Christen M., Bernstein A. Implementations in Machine Ethics: A Survey. *ACM Computing Surveys*. Vol. 53, Is. 6. Art. 132. 2021. P. 1–38. DOI: 10.1145/3419633.
2. Taylor J., Deng W.H., Holstein K., Fox S., Zhu H. Carefully Unmaking the “Marginalized User” // *ACM Transactions on Computer-Human Interaction*. 2024. Vol. 31, Is. 6. Art. 81. P. 1–30. DOI: 10.1145/3673229.
3. Лукичев П.М., Чекмарев О.П. Применение искусственного интеллекта в системе высшего образования // *Вопросы инновационной экономики*. 2023. Т. 13. № 1. С. 485–502. DOI: 10.18334/vines.13.1.117223.
4. Bennett S.J. Transmuting values in artificial intelligence: investigating the motivations and contextual constraints shaping the ethics of artificial intelligence practitioners. *Science and Technology Studies*. Thesis or Dissertation. Doctoral. 2023. Edinburgh College of Art thesis and dissertation collection. [Электронный ресурс]. URL: <https://hdl.handle.net/1842/41155> (дата обращения: 05.07.2025). DOI: 10.7488/era/3891.
5. Owe A., Baum S.D. Moral consideration of nonhumans in the ethics of artificial intelligence. *AI Ethics* 1. 2021. P. 517–528. DOI: 10.1007/s43681-021-00065-0.
6. Ashok M., Madan R., Joha A., Sivarajah U. Ethical framework for Artificial Intelligence and Digital technologies // *International Journal of Information Management*. 2022. Vol. 62. P. 1–62. 102433. DOI: 10.1016/j.ijinfomgt.2021.102433.
7. Ibo van de Poel. Embedding Values in Artificial Intelligence (AI) Systems // *Minds and Machines*. 2020. Vol. 30. P. 385–409. DOI: 10.1007/s11023-020-09537-4.
8. Schenk P., Müller V., Keiser L. Social Status and the Moral Acceptance of Artificial Intelligence // *Sociological Science*. 2024. Vol. 11, Is. 10. P. 989–1016. DOI: 10.15195/v11.a36.
9. Hastuti R.S. Ethical Considerations in the Age of Artificial Intelligence: Balancing Innovation and Social Values // *West Science Social and Humanities Studies*. 2023. Vol. 01, Is. 2. P. 78–87. DOI: 10.58812/wsshs.v1i02.191.
10. Holmes W., Porayska-Pomsta K., Holstein K., Sutherland E., Baker T., Shum S.B., Santos O.C., Rodrigo M.T., Cukurova M., Bittencourt I.I., Koedinger K.R. Ethics of AI in education:

Towards a community-wide framework // *International Journal of Artificial Intelligence in Education*. 2022. № 32 (3). P. 504–526. DOI: 10.1007/s40593-021-00239-1.

11. Borenstein J., Howard A. Emerging challenges in AI and the need for AI ethics education // *AI and Ethics*. 2021. № 1 (1). P. 61–65. DOI: 10.1007/s43681-020-00002-7.

12. Daradoumis T., Arguedas M. Cultivating students' reflective learning in metacognitive activities through an affective pedagogical agent // *Educational Technology and Society*. 2020. № 23 (2). P. 19–31.; URL: <https://www.jstor.org/stable/26921131> (дата обращения: 05.07.2025).

13. Kazim E., Koshiyama A.S. A high-level overview of AI ethics // *Patterns*. 2021. № 2 (9). 100314. DOI: 10.1016/j.patter.2021.100314.

14. Nigam A., Pasricha R., Singh T., Churi P. A systematic review on ai-based proctoring systems: Past, present and future // *Education and Information Technologies*. 2021. № 26 (5). P. 6421–6445. DOI: 10.1007/s10639-021-10597-x.

15. Reiss M.J. The use of AI in education: Practicalities and ethical considerations // *London Review of Education*. 2021. № 19 (1). P. 1–14. DOI: 10.14324/LRE.19.1.05.

16. Bonnefon J.-F., Rahwan I., Shariff A. The Moral Psychology of Artificial Intelligence // *Annual Review of Psychology*. 2024. Vol. 75, Is. 1. P. 653–675. DOI: 10.1146/annurev-psych-030123-113559.

17. Telkamp J.B., Anderson M.H. The Implications of Diverse Human Moral Foundations for Assessing the Ethicality of Artificial Intelligence // *J Bus Ethics*. 2022. Vol. 178. P. 961–976. DOI: 10.1007/s10551-022-05057-6.

18. Аузан А.А., Бахтигараева А.И., Брызгалин В.А., Золотов А.В., Никишина Е.Н., Припузова Н.А., Ставинская А.А. Социокультурные факторы в экономике: пройденные рубежи и актуальная повестка // *Вопросы экономики*. 2020. № 7. С. 75–91. DOI: 10.32609/0042-8736-2020-7-75-91. EDN: YLTOWI.