

**ГАРМС Александр Александрович,**  
Магистрант, Московский государственный технический университет  
имени Н.Э. Баумана,  
garms.alexander@yandex.ru

**Научный руководитель:**  
**Оплетина Надежда ВИТАЛЬЕВНА**  
кандидат социологических наук, доцент кафедры «Социология  
и культурология» МГТУ им. Н.Э. Баумана,  
opletinav@bmstu.ru

**Alexander Alexandrovich GARMS,**  
Master's Student of the Bauman Moscow State Technical University,  
garms.alexander@yandex.ru

**Scientific adviser:**  
**Nadezda Vitalievna OPLETINA,**  
Ph. D. (Sociological), Associate Professor of the Department of Sociology  
and Cultural Studies, Bauman Moscow State Technical University,  
opletinav@bmstu.ru

УДК: 316.4

# МЕЖДУ ЭТИКОЙ И ПРОГРЕССОМ: проблемы развития искусственного интеллекта

## BETWEEN ETHICS AND PROGRESS: problems of development of artificial intelligence

### Аннотация

Анализируется процесс внедрения технологий искусственного интеллекта в жизнь современного общества. Рассмотрен амбивалентный характер влияния искусственного интеллекта на общественную жизнь. Показано, что в условиях стремительного развития технологий искусственного интеллекта моральные, социальные и культурные аспекты становятся критически важными для получения релевантных и точных оценок социальных практик использования инструментов искусственного интеллекта. При этом стремительное развитие ИИ создает значительные этические проблемы и продуцирует социальные риски.

**Ключевые слова:** технологии ИИ, искусственный интеллект, этическая дилемма, парадокс доверия, этические риски, этика искусственного интеллекта.

### Abstract

This article analyzes the process of introducing artificial intelligence technologies into the life of modern society. The ambivalent nature of the influence of artificial intelligence on public life is considered. It is shown that with the rapid development of artificial intelligence technologies, moral, social and cultural aspects become critical for obtaining relevant and accurate assessments of social practices in the use of artificial intelligence tools. At the same time, the rapid development of AI creates significant ethical problems and generates social risks.

**Keywords:** AI technologies, artificial intelligence, ethical dilemma, paradox of trust, ethical risks, ethics of artificial intelligence.

Последний год стал прорывным в развитии искусственного интеллекта. В интернет-пространстве развернулось своеобразное соревнование, кто еще какую задачу автоматизирует с помощью ChatGPT, и какой фейк от Midjourney лучше «завирусится», а техно-миллионеры, типа Илона Маска, активно инвестируют в создание «настоящего» ИИ – такого, который сможет сам учиться, развиваться и решать любые задачи, и даже такие, которые человечество не умело или не смогло решать раньше.

По прогнозам Всемирного экономического форума, в ближайшем будущем огромное количество людей будут напрямую работать с технологиями искусственного интеллекта. И уже к 2025 году для удовлетворения потребностей индустрии искусственного интеллекта потребуется около 97 миллионов специалистов, не считая того факта, что внедрение ИИ-ботов, автоматизации и подобных инструментов, безусловно, повлияет на бесчисленное количество будущих рабочих мест [24].

Сегодня эксперты прогнозируют, что наибольшие экономические выгоды от ИИ будут получены в Китае (увеличение ВВП на 26% в 2030 году) и Северной Америке (увеличение на 14,5%), что в сумме составит 10 трлн долларов [23]. Именно китайские и индийские компании чаще всего повсеместно внедряют используют ИИ в (почти 60%), за ними следуют США (25%), Великобритания (26%), Канада (28%), Австралия (24%) [17].

Современные достижения в области робототехники и развития искусственного интеллекта показывают, что высокотехнологичные системы становятся повсеместными и повседневными, и все более способными, и независимыми. В то же время, автономные интеллектуальные системы допускают ошибки при принятии решений, а их функционирование может привести человеческим жертвам. Среди ученых сформировалось и активно распространяется в информационном пространстве мнение о серьезных рисках для общества и человечества в связи с сверхинтенсивным развитием технологий ИИ.

Так, известный американский ученый, специалист по ИИ, Элиезер Юджовский заявляет об угрозе жизни всех людей на Земле из-

за бесконтрольных разработок искусственного интеллекта в последние годы [28]. А в марте 2023 года более тысячи ученых (в том числе Илон Маск, Стив Возняк и другие ведущие специалисты в сфере ИИ) опубликовали открытое письмо, призывающее к немедленной приостановке гигантских экспериментов с искусственным интеллектом, таких как ChatGPT (генеративный чат-бот с ИИ, способный к осмысленному диалогу с человеком) [25].

Данные обстоятельства способствуют тому, что все больше и больше людей включаются в публичное обсуждение нарративов этики искусственного интеллекта (ИИ), которые все более фокусируются на морально-этических дилеммах использования ИИ в жизненно важных ситуациях, на социальных рисках применения ИИ в повседневной жизни, когда нарушается приватность, увеличивается непрозрачность применения таких технологий, а так же нарушается справедливость, при их использовании в широкой практике. Например, массовое внедрение роботов по-прежнему вызывает серьезные опасения, в том числе моральное неприятие лишения человека жизни в результате функционирования интеллектуальной автоматической системы из-за ошибки машины. Активное распространение технологий ИИ актуализирует в общественном сознании так называемую «проблему вагонетки» – этическую дилемму, исследуемую в когнитивистике и нейрорэтике, и связанную с моральным выбором при принятии этических решений. Исследователи и политики спорят о том, нужно ли перекладывать ответственность за человеческие жизни на алгоритмы и способна ли бесчувственная машина на моральный выбор.

В то же время, любое этическое поведение, исследующее такие вопросы, должно учитывать различные контексты, в которых происходит взаимодействие интеллектуальных систем и общества. При этом, учет морально значимых, социальных и культурных аспектов, является критически важным для получения релевантных и точных оценок социальных практик использования инструментов искусственного интеллекта. Более того, проблема усугубляется тем фактом, что не всегда возможно логично объяснить процес-

сы, связанные с этическими решениями, чтобы «научить» им машин и сделать ее способной проявлять этическую разборчивость, особенно когда машины становятся активными участниками общественных процессов, которые до сих пор являлись областью деятельности исключительно человека.

Таким образом, важной задачей становится решение вопросов понимания сложной взаимосвязи между положительными и отрицательными сторонами ИИ, что предполагает критический анализ компромиссов, присущих ландшафту ИИ, и необходимость ответственного подхода к разработке и внедрению ИИ. Как мы можем смягчить негативные последствия и одновременно усилить положительный эффект ИИ? Какие этические рамки, нормы и гарантии необходимы для обеспечения соответствия ИИ человеческим ценностям и устремлениям? Решение этих вопросов требует всестороннего понимания сложной и развивающейся взаимосвязи между ИИ и жизнью общества с учетом как его технологических достижений, так и социально-культурных, экономических и политических последствий.

Многие ведущие эксперты отмечают прежде всего положительные моменты внедрения технологий искусственного интеллекта в жизнь общества, которые, по их мнению, связаны с повышением уровня жизни человека в условиях масштабирования технологий искусственного интеллекта. Так, Мартин ван Оттерло, автор книги «Gatekeeping Algorithms with Human Ethical Bias» и доцент кафедры искусственного интеллекта Тилбургского университета (Нидерланды), пишет: «Несмотря на то, что я вижу множество этических проблем, потенциальных проблем и особенно проблем дисбаланса власти/неправильного использования ИИ (не говоря уже о проблемах сингулярности и неконтролируемого ИИ), я считаю, что ИИ изменит жизнь большинства людей к лучшему, особенно если смотреть на ближайшую перспективу до 2030 года, тем более что даже плохие последствия ИИ могут рассматриваться большинством людей как преимущественно «хорошие» [29].

Действительно, ИИ способен значительно улучшить жизнь человека в различных областях, таких как здравоохранение и образо-

вание; технологии ИИ постепенно вытесняют многие физически трудные или рутинные виды работы; активно используются возможности искусственного интеллекта и для решения глобальных проблем. Уже сегодня искусственный интеллект позволяет оптимизировать производственные процессы, персонализировать обучение и предлагает инновационные решения сложных задач, что позволяет утверждать, что ответственное внедрение и использование технологий ИИ может привести к значительным достижениям и выгодам для человечества.

В медицине уже используется большое количество технологий искусственного интеллекта, помогающих медикам оптимизировать клинические и административные процессы в здравоохранении. Так, в 2018 году было задействовано более 5 тысяч медицинских роботов, которые ассистировали хирургам более чем в 1 млн операций различной степени сложности. В качестве примера успешной работы роботом-ассистентов можно привести итоги исследования, в котором приняли участие пациенты, перенесшие ортопедические вмешательства. Исследование показало, что использование роботов с искусственным интеллектом в качестве ассистентов привело к пятикратному снижению осложнений по сравнению с операциями, выполняемыми хирургами в одиночку [18].

Системы здравоохранения на основе ИИ интенсивно развиваются: за период с 2005 по 2020 гг. в мире количество случаев внедрения ИИ в медицинские процессы выросло почти в 62 раза, особенно в области раннего обнаружения и диагностики. Эти разработки позволяют ИИ выполнять задачи, которые иногда не под силу человеку, при этом обеспечивая хорошую эффективность решения задач, простоту, надежность и тщательность при меньших затратах. В целом современный вектор развития ИИ в здравоохранении показывает, что данные технологии могут значительно повысить качество обслуживания пациентов и одновременно снизить расходы на здравоохранение [9].

Специалисты видят перспективы широкого внедрения ИИ и в повседневную жизнь человека и одним из самых заметных проявлений ИИ в повседневной жизни является использование его в домашних устройствах. Голосовые

помощники, такие как Siri, Google Assistant, Яндекс Алиса и Amazon Alexa, позволяют находить информацию, управлять домашней техникой, заказывать товары, и др. Современные домашние устройства благодаря ИИ стали более интуитивными и удобными в использовании. По данным исследования Juniper Research, в 2023 году в мире будет использоваться более 8 миллиардов голосовых помощников. Для сравнения, на конец 2018 года голосовых помощников было 2,5 миллиарда [26].

Еще одним преимуществом технологий ИИ выступает тот факт, что с помощью алгоритмов искусственного интеллекта возможно легко справляться даже с утомительной рутинной работой. При этом искусственный интеллект может работать бесконечно, без перерывов, он думает гораздо быстрее людей и выполняет несколько задач одновременно с точным результатом. Примером могут служить чат-боты для онлайн-поддержки клиентов, которые могут оказывать мгновенную помощь клиентам в любое время и в любом месте. Используя искусственный интеллект и обработку естественного языка, чат-боты могут отвечать на общие вопросы, решать проблемы и переадресовывать сложные задачи агентам, обеспечивая бесперебойное обслуживание клиентов в круглосуточном режиме [10].

Сегодня искусственный интеллект успешно зарекомендовал себя как преобразующая сила в самых разных отраслях, в том числе и в образовании. 99,4% из 509 высших учебных заведений США заявили, что искусственный интеллект в образовании и обучении будет играть важную роль в конкурентоспособности их учебного заведения в ближайшие три года. В мире, где технологии развиваются беспрецедентными темпами, нет ничего удивительного в том, что искусственный интеллект проникает в учебные аудитории. По мере того, как преподаватели и исследователи продолжают изучать возможности ИИ в образовании, они обнаруживают его потенциал, способный революционизировать процесс обучения. По прогнозам, к 2027 году объем рынка ИИ в образовании превысит 20 млрд. долларов [22].

Образование, как ведущий компонент современного общественного развития эво-

люции общества и личности, безусловно, способно получить значительные преимущества от внедрения ИИ. Интеграция искусственного интеллекта в образовательные системы меняет подходы к обучению студентов, преподаванию учителей и функционированию учебных заведений. Благодаря персонализации учебного процесса, автоматизации административных функций и обратной связи в режиме реального времени ИИ революционизирует образовательный ландшафт, устраняет пробелы и способствует формированию более инклюзивной и эффективной среды обучения.

Аналитики утверждают, что потенциальные возможности применения ИИ в образовании включает в себя персонализированное обучение, интеллектуальные системы обучения, автоматизацию оценивания и взаимодействия между учителем и учеником, студентом и преподавателем, а также могут помочь улучшить результаты обучения и его эффективность. Алгоритмы ИИ, такие как обучение с подкреплением, могут быть использованы для динамического изучения индивидуальных потребностей студента и соответствующей адаптации процесса обучения. В связи с персонализированным обучением могут быть разработаны интеллектуальные обучающие системы, способные активно взаимодействовать с учащимися, давая им ценную обратную связь. А масштабируемость ИИ позволит высококачественному образованию стать доступным, обеспечивая повсеместный доступ к нему [19].

Высокие темпы внедрения и распространения решений с использованием технологий машинного обучения мотивируют общество привыкать к их повсеместному использованию и научиться доверять таким системам. В то же время, проводимые исследователями замеры свидетельствуют, о некотором этической дилемме, проявляющейся в общественном мнении как «парадокс доверия» интеллектуальным системам, которые, активно участвуя в трансформации социальных практик и процессах принятия решений, становятся «новыми» посредниками нашего существования [3]. В совместном исследовании международной аудиторской фирмы KPMG и Университета Квинсленда (Австралия), посвященном теме доверия

к искусственному интеллекту (ИИ) зафиксировано, что люди не доверяют ИИ, но допускают его использование: большинство опрошенных указало, что относятся терпимо к ИИ (42%) или нейтрально (28%). Меньшинство опрошенных одобряют использование ИИ (15%), поддерживают (6%). 9% опрошенных относятся к ИИ негативно [14].

В России аналогичное исследование провел ВЦИОМ: «Искусственный интеллект: угроза или светлое будущее?» Россияне ощущают расширение применения технологий ИИ и, в первую очередь, это касается систем видеонаблюдения (79%) и робототехники (74%) Более половины опрошенных также полагают, что технологии искусственного интеллекта используются в приложениях для обработки фотографий (62%), для онлайн-заказов (61%) и вызова такси (59%), в медицине (56%), в службах поддержки клиентов и логистике (55%) и в финансовой сфере (50%). В целом, российское общество демонстрирует положительное отношение к использованию технологий ИИ. Больше всего россияне одобряют внедрение систем ИИ в промышленность (82%), науку (79%) и строительство (76%) [2].

Таким образом, парадокс ситуации, сложившейся в общественном мнении в том, что люди понимают неизбежность внедрения и распространения систем искусственного интеллекта, но не готовы доверять этим системам, т.к. уверены в том, что данные системы не эмоциональны и аморальны, не могут опираться на морально-этические нормы при выстраивании взаимодействия. Данные интеллектуальные системы будут осуществлять выбор в обязательном порядке и при этом, какое бы решение не было принято, оно будет противоречить нормам морали.

Например, все тот же опрос ВЦИОМ зафиксировал, что, несмотря на высокий уровень одобрения, россияне пока не готовы доверить искусственному интеллекту принятие конечных решений в той или иной сфере. При оказании медицинских и образовательных услуг использование специалистами ИИ допустимо в случае, если решения принимает человек, а искусственный интеллект только помо-

гает, – так считают 74% и 62% соответственно. Более половины россиян выступают против замены водителя искусственным интеллектом при управлении транспортными средствами на дорогах общего пользования (59%). Примерно такая же доля опрошенных допускают использование технологии распознавания лиц в целях обеспечения правопорядка лишь при условии сохранения личных данных и прав граждан на личную жизнь (57%) [14].

Риски, связанные с технологиями ИИ при взаимодействии с человеком, безусловно, включают в себя этические и моральные проблемы, которые могут продуцироваться в результате ошибок, вызванных данными или алгоритмами. В качестве примеров этических рисков, связанных с технологиями искусственного интеллекта можно привести выбор интеллектуальной системы между жизнью пешеходов и водителей в случае аварийной ситуации, нарушение прав на частную жизнь людей, участвующих в «поиске человеческой плоти» (китайский термин, обозначающий феномен определенных исследований с использованием интернет-СМИ, таких как блоги и форумы) на основе технологии больших данных, а также неверные решения, принимаемые «интеллектуальными судами», лишенными человеческих эмоций и чувств. ИИ часто не справляется со сложными сценариями принятия решений, поскольку неявные знания, такие как обычаи, эмоции и убеждения, трудно полностью оцифровать и структурировать. В то же время вопрос о том, сможет ли будущий интеллектуальный процесс принятия решений в эпоху сильного ИИ превзойти или даже заменить человеческий выбор, представляет собой прежде всего «моральную дилемму».

И как было отмечено выше, общественное мнение интуитивно улавливает данные противоречия, определяя технологии ИИ как опасные и неопределенные. Ведь принятие решений в области искусственного интеллекта основано на ограниченных данных, программах, соответствующих алгоритмах и других входных условиях для разработки возможной стратегии. Сама же технология ИИ сопряжена с неопределенностью, и в сочетании с неполным характером данных, решений, в которых

отсутствует человеческая эмпатия, как правило, подвержены ошибкам. Это влечет за собой этические риски, отражение которых мы и находим в отношении людей к технологиям ИИ: нарушение конфиденциальности, опасность для жизни человека, подрыв социальной справедливости – данные неопределенности являются важным источником возникающих этических дилемм.

Об этом же говорят и эксперты, занимающиеся разработкой и внедрением технологий искусственного интеллекта, утверждая, что источники этических рисков при принятии решений ИИ включают две основные причины: технологическую неопределенность и ограниченную рациональность человека [15].

В первом случае, неправильное использование технологий и злоупотребления ими являются самыми большими источниками технологического риска [12]. В частности, исследователи отмечают, что интеллектуальные алгоритмы, разработка программ и другие технологии, которые существуют во всем процессе принятия решений искусственным интеллектом, являются специфическими источниками этических рисков [21]. Во-втором случае, ограниченная рациональность человека, описанная в работах Д. Канемана, П. Словика, А. Тверски, становится самостоятельным источником этических рисков, поскольку образцы программирования и импорта данных при интеллектуальном принятии решений связаны с человеческими решениями [1], а этические риски при принятии решений ИИ возникают из-за сложных взаимодействий между технологиями, людьми, обществом и природой.

Системы искусственного интеллекта обучаются на огромных массивах данных, и в этих данных заложены общественные предубеждения. Следовательно, эти предубеждения могут укорениться и в алгоритмах ИИ, закрепляя и усиливая несправедливые или дискриминационные результаты в таких важнейших областях, как прием на работу, кредитование, уголовное правосудие и распределение ресурсов. Например, если компания использует систему искусственного интеллекта для отбора кандидатов на работу путем анализа их резюме, то эта система, скорее всего, была обуче-

на на имеющемся опыте и соответствующих данных об успешных приемах на работу в компании. Однако если данные являются необъективными, например содержат гендерные или расовые предубеждения, то система искусственного интеллекта также способна усвоить эти предубеждения, дискриминируя кандидатов, не соответствующих имеющимся в компании данным о найме. Несколько американских ведомств недавно опубликовали свои предложения о том, как они намерены бороться с предвзятостью в моделях искусственного интеллекта и привлекать организации к ответственности за укоренение дискриминации с помощью своих платформ [13].

Еще одним опасением, активно продвигаемым и обсуждаемым в публичном информационном пространстве на протяжении последних лет является потенциальная угроза того, что технологии искусственного интеллекта и автоматизации способны вытеснить людей из большинства отраслей экономики. Действительно, по оценкам специалистов, роботизированное производство окупается уже через девятнадцать месяцев и в целом повышает эффективность и производительность. Сегодня 71% рабочего времени относятся к человеческим трудозатратам, и лишь 29% берут на себя машины. Но уже к 2025 году, по оценкам специалистов, этот показатель может измениться до 48% и 52% соответственно, то есть большую часть работы начнут выполнять машины [11].

А согласно отчету Всемирного экономического форума The Future of the Jobs 2018, уже к 2030 году 75 миллионов рабочих мест по всему миру будут упразднены, благодаря внедрению интеллектуальных технологий, в то же время их заменят другие 133 млн. Но готово ли трудоспособное человечество к тому, чтобы занять данные места, каким образом необходимо формировать компетенции для таких мест и способно ли современное образование к эффективному ответу на данные вызовы. При этом эксперты отмечают, что общественный страх того, что «роботы заменят людей», не соответствует реальности, точнее будет формулировка «упростят и повысят эффективность работы» [21].

С этим же вызовом столкнулись и большинство современных компаний, перед которыми встает множество вопросов, связанных с адаптацией сотрудников к новым интеллектуальным технологиям. А решение вопроса, как всегда, лежит в этической плоскости: кого-то придется заменить, кому-то необходимо повышать свои компетенции или полностью менять специальность. При этом, далеко не все работодатели готовы обучать действующих сотрудников, дешевле – нанимать новых. Например, в России агентство CNews Analytics провело исследование, целью которого было определить, какие технологии будут наиболее востребованы отечественным бизнесом в ближайшее время. Согласно результатам, наибольшие надежды в 2022 г. возлагаются на аналитику больших данных, искусственный интеллект и облачные решения. В компаниях реального сектора, по мнению опрошенных CNews Analytics, будут востребованы автономные системы и решения в сфере интернета вещей.

Для снижения рисков в условиях столь масштабных преобразований в сфере труда только в рамках федерального проекта «Кадры для цифровой экономики» национальной программы «Цифровая экономика» организовано качественное обучение молодых людей и переобучение возрастных специалистов, что позволит им адаптироваться под новые требования рынка, освоить IT-компетенции и сформировать необходимые навыки [5]. А Минэкономразвития открыло набор слушателей на курсы по обучению в области искусственного интеллекта, в рамках обучения государство возместит до 85% стоимости курсов. Объем финансирования программы на ближайшие три года составит 575 млн рублей [4].

Вместе с тем, и в настоящее время именно потеря рабочего места или неудачная попытка устроиться на более высокооплачиваемую работу вызывает наибольшие опасения у 33% россиян, свидетельствуют данные опросов [6]. Так же аналитики выяснили, что в основном опрошенные боятся сокращения рабочих из-за развития технологии ИИ. Так, боятся, что искусственный интеллект может занять их рабочие места, 58%

респондентов. В том, что их профессии могут полностью исчезнуть из-за технологии, уверены 11% опрошенных, а в том, что ИИ сможет частично перенять их функционал, не сомневаются 46% из них [8].

По мере расширения использования искусственного интеллекта, особое значение приобретает подход к внедрению ИИ, ориентированный на человека. Человеко-ориентированный ИИ способен создать определенный баланс: чем больше людей сотрудничают с ИИ, тем лучше будут результаты. А если результаты будут лучше, то больше людей будут использовать ИИ и доверять ему.

В то же время, сегодня можно диагностировать определенное статусное неравенство использования ИИ современными организационными системами. Например, исследование, проведенное Лабораторией инноваций в сфере труда, показывает, что руководители компаний лидируют по внедрению ИИ: более половины (52%) используют ИИ на еженедельной основе, в то время как среди остальных сотрудников этот показатель составляет всего 36% [16]. Это свидетельствует о том, что руководители лучше понимают, как ИИ может улучшить процесс принятия решений и способствовать инновациям, однако другие уровни ролей отстают.

Если рассматривать ИИ как партнера, а не как замену творческого потенциала человека, то возможно побудить еще большее число сотрудников принять и доверять ИИ. По данным исследования, сотрудники уже находят практические способы взаимодействия с ИИ: 30% сотрудников используют интеллектуальные технологии для анализа данных с целью формирования практических выводов, а 25% сотрудников – для решения административных задач, освобождая время для более стратегической работы. Таким образом, а подход, ориентированный на человека, может помочь расширить использование ИИ и повысить доверие к нему [16].

В целом, современные исследования свидетельствуют, что для того, чтобы ИИ стал основным элементом совместной деятельности, необходим человеко-ориентированный подход. Взаимодействие между людьми и ис-

кусственным интеллектом позволяет достичь лучших результатов и способствует росту доверия и дальнейшему использованию искусственного интеллекта.

Стоит признать, что людям сегодня необходим сбалансированный и взвешенный подход к оценке технологий ИИ. Для получения максимальных выгод от ИИ и снижения рисков необходимо всестороннее понимание его последствий. Это включает в себя рассмотрение прежде всего этических основ, прозрачности алгоритмов, подотчетности и сотрудничества между заинтересованными сторонами для обеспечения соответствия технологий искусственного интеллекта человеческим ценностям и устремлениям.

Понимание данного момента уже оформилось в публичном пространстве. «Государствам мира необходимо придерживаться сбалансированного подхода к искусственному интеллекту и не вводить нормативные акты, которые будут препятствовать инновациям» – заявил А. Кант, председатель G20, выступая на саммите SAI20 13 марта 2023 года в Гоа [10]. Он подчеркнул, что стремительное развитие таких технологий, как генеративный искусственный интеллект, открывает новые возможности для развития человечества, но также несет в себе новые риски (например, фальшивые новости), новые формы дискриминации и предвзятости, угрозы личной жизни и защите данных. «К этой технологии (ИИ) должен быть взвешенный подход. Такой, который позволил бы использовать все преимущества ИИ и одновременно минимизировать вред, который он может нанести, поскольку технологии всегда будут опережать правительство, и чем больше мы будем регулировать, тем сильнее будет проблема остановки инноваций», – сказал Кант [10].

В государственном университете признали, что, хотя ИИ, обычно определяемый как имитация человеческого интеллекта машинами, обладает огромным потенциалом для «преобразования образования» и уже «облегчает нашу жизнь», автоматизируя все больший перечень задач, он также связан с серьезными рисками и проблемами, с которыми необходимо бороться и решать.

16 ноября в пресс-центре ТАСС состоялась «Форум этики в сфере искусственного интеллекта: Поколение GPT. Красные линии ИИ». Мероприятие стало площадкой для демонстрации лучших практик применения этики искусственного интеллекта, присоединения новых организаций к Кодексу этики в сфере ИИ и формирования актуального комплекса проектов по совершенствованию этического и нормативного регулирования ИИ. Ученые, разработчики прикладных ИИ-решений, представители сферы образования, здравоохранения, культуры и искусства выразили свои мысли о этических вопросах, связанных с широким использованием интеллектуальных технологий. Они обсуждали роль государства в правовом регулировании области машинного обучения и его влияния на культуру, образование и автоматизацию производства. Эксперты выразили опасения относительно возможной деградации сферы знаний и образования, а также о прямом негативном влиянии на население при делегировании многих процессов и действий технологиям. Спикеры пришли к выводу, что использование технологий искусственного интеллекта должно быть основано на принципах права, морали и безопасности, и что цифровые решения следует воспринимать не как замену человека, а как помощь в его деятельности [7].

Таким образом, сегодня искусственный интеллект стал неотъемлемой частью многих отраслей жизни современного общества, кардинально изменив методы ведения бизнеса и внедрения инноваций, да и повседневную жизнь людей. При этом, стремительное развитие ИИ создает значительные этические проблемы и риски для жизни человека, что показывают многие исследования доверия к технологиям ИИ в мире. И сегодня, без надлежащего регулирования прежде всего этических рамок ИИ, мир может столкнуться с масштабными рисками внедрения и использования технологий искусственного интеллекта. А разработка ясных этических ориентиров, позволят предотвратить или минимизировать многие ошибки, способные повлечь за собой неблагоприятные для человека последствия и, возможно прийти к решению этической дилеммы, связанной с конфронтацией норм и ценностей доверия к технологиям ИИ.

## ЛИТЕРАТУРА

1. Аркин Р.С. Управление летальным поведением: внедрение этики в гибридную совещательную / реактивную архитектуру робота // Материалы 3-й международной конференции ACM / IEEE по взаимодействию человека и робота. 2008. С. 121–128.
2. Искусственный интеллект: угроза или светлое будущее? 28 декабря 2022 г. / Сайт ВЦИОМ. [Электронный ресурс]. – URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/iskusstvennyi-intellekt-ugroza-ili-svetloe-budushchee> (дата обращения: 18.11.2023).
3. Лешкевич Т.Г. Парадокс доверия к искусственному интеллекту и его обоснование // Философия науки и техники. 2023. Т. 28. №. 1. С. 34–47.
4. Минэкономразвития открыло набор на курсы в области ИИ с возмещением до 85% их стоимости. 8 августа 2022 г. / Сайт ТАСС. [Электронный ресурс]. – URL: <https://tass.ru/ekonomika/15419519> (дата обращения: 23.11.2023).
5. Петрова О.П. Тенденции развития российского рынка ИТ-услуг в условиях цифровой экономики / О.П. Петрова // Цифровая экономика, информационное общество и информационная безопасность: основные социально-экономические аспекты: Материалы международной научно-практической конференции, Смоленск, 29 марта 2023 года. Смоленск: Издательство «Маджента», 2023. С. 282–287. EDN VDBHVZ.
6. Социологический опрос Исследовательского центра портала Superjob.ru. 5 января 2021 г. / Сайт Superjob.ru. [Электронный ресурс]. – URL: <https://www.interfax-russia.ru/moscow/news/tret-rossiyan-opasayutsya-poteri-raboty-v-2021-godu-opros> (дата обращения: 23.11.2023).
7. Форум этики в сфере искусственного интеллекта «Поколение GPT. Красные линии» 16 ноября 2023 года [Электронный ресурс]. – URL: <https://aiethic.ru/> (дата обращения: 26.11.2023).
8. Мамиконян О. 58% россиян боятся сокращения рабочих мест из-за развития искусственного интеллекта. / Сайт Forbes. [Электронный ресурс]. – URL: <https://www.forbes.ru/forbeslife/481810-58-rossiyan-boatsa-sokrasenia-rabocih-mest-iz-za-razvitia-iskusstvennogo-intellekta> (дата обращения: 23.11.2023).
9. Ali O. et al. A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge*, 2023, vol. 8, no. 1, pp. 100333.
10. Avundukluoplu P. SAI20 2023 Gьndemi: Mavi Ekonomi ve Sorumlu Yapay Zeka. Yьl 2023, Sayь: 128, 169–176.
11. Bhbosale S., Pujari V., Multani Z. Advantages and disadvantages of artificial intelligence. *Aayushi International Interdisciplinary Research Journal*, 2020, vol. 77, pp. 227–230.
12. Cann O. Machines will do more tasks than humans by 2025 but robot revolution will still create 58 million net new jobs in next five years. *World Economic Forum*, 2018, vol. 17.
13. Chen X.P. The target, tasks, and implementation of artificial intelligence ethics: Six issues and the rationale behind them. *Philos. Res*, 2020, vol. 9, pp. 79–87.
14. Chen Z. Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 2023, vol. 10, no. 1, pp. 1–12.
15. Gillespie N., Lockey S., Curtis C. Trust in artificial intelligence: A five country study. – 2021.
16. Guan H., Dong L., Zhao A. Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making. *Behavioral Sciences*, 2022, vol. 12, no. 9, p. 343.
17. Human-centric AI at work: A playbook for powering your organization with AI. *The Work Innovation Lab*, 2023.
18. Index I.B.M. Global AI Adoption Index IBM Corporation. 2022.
19. Kalis B., Collier M., Fu R. 10 promising AI applications in health care. *Harvard business review*, 2018.
20. Kamalov F., Santandreu Calonge D., Gurrib I. New era of artificial intelligence in education: towards a sustainable multifaceted revolution. *Sustainability*, 2023, vol. 15, no. 16, p. 12451.
21. Marabelli M., Newell S., Handunge V. The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, 2021, vol. 30, no. 3, p. 101683.
22. Opletina N., Kunyaeva M. Technology Assessment as a field of professional competency of technoscience society engineers. *2022 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2022. P. 41–44.
23. Pham S.T.H., Sampson P.M. The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning*, 2022, vol. 38, no. 5, pp. 1408–1421.
24. Raporu PWC Yapay Zeka. PwC’s Global Artificial Intelligence Study: Exploiting the AI Revolution // Bkz. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html> (дата обращения: 22.10.2023).

25. *Russo A.* Recession and automation changes our future of work, but there are jobs coming, report says. *World Econ Forum [Online]*, 2020.
26. *Samuel J.* Response to the March 2023 'Pause Giant AI experiments: an open letter' by Yoshua Bengio, signed by Stuart Russell, Elon Musk, Steve Wozniak, Yuval Noah Harari and others... *Elon Musk, Steve Wozniak, Yuval Noah Harari and others...* (March 29, 2023). 2023.
27. *Smith S.* Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices. *Juniper Research*, 2018.
28. *Turchin A.* Structure of the global catastrophe. Risks of human extinction in the XXI century. *Lulu.com*, 2010, no. 09, pp. 67–69.
29. Van Otterlo M. Gatekeeping algorithms with human ethical bias: The ethics of algorithms in archives, libraries and society // arXiv preprint arXiv:1801.01705. 2018.

## REFERENCES

1. *Arkin R.S.* Upravlenie letal'nym povedeniem: vnedrenie jetiki v gibridnuju soveshhatel'nuju. *Reaktivnuju arhitekturu robota. Materialy 3-j mezhdunarodnoj konferencii ACM. IEEE po vzaimodejstvu cheloveka i robota*, 2008, pp. 121–128.
2. Iskusstvennyj intellekt: ugroza ili svetloe budushchee? 28 dekabrya 2022 g. Sajt VCIOM. [Jelektronnyj resurs]. – URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/iskusstvennyi-intellekt-ugroza-ili-svetloe-budushchee> (data obrashhenija: 18.11.2023).
3. *Leshkevich T.G.* Paradoks doverija k iskusstvennomu intellektu i ego obosnovanie. *Filosofija nauki i tehniki*, 2023, vol. 28, no 1, pp. 34–47.
4. Minjekonomrazvitija otkrylo nabor na kursy v oblasti IT s vozmeshheniem do 85% ih stoimosti. 8 avgusta 2022 g. Sajt TASS. [Jelektronnyj resurs]. – URL: <https://tass.ru/ekonomika/15419519> (data obrashhenija: 23.11.2023).
5. *Petrova O.P.* Tendencii razvitija rossijskogo rynka IT-uslug v uslovijah cifrovoj jekonomiki. *Cifrovaja jekonomika, informacionnoe obshhestvo i informacionnaja bezopasnost': osnovnye social'no-jekonomicheskie aspekty: Materialy mezhdunarodnoj nauchno-prakticheskoy konferencii, Smolensk, 29 marta 2023 goda*. Smolensk: Izdatel'stvo «Madzhenta», 2023. Pp. 282–287. EDN VDBHVZ.
6. Sociologicheskij opros Issledovatel'skogo centr portala Superjob.ru. 5 janvarja 2021 g. / Sajt Superjob.ru. [Jelektronnyj resurs]. – URL: <https://www.interfax-russia.ru/moscow/news/tret-rossiyan-opasayutsya-poteri-raboty-v-2021-godu-opros> (data obrashhenija: 23.11.2023).
7. Forum jetiki v sfere iskusstvennogo intellekta «Pokolenie GPT. Krasnye linii» 16 nojabrya 2023 goda [Jelektronnyj resurs]. – URL: <https://aiethic.ru/> (data obrashhenija: 26.11.2023).
8. *Mamikonjan O.* 58% rossiyan bojatsja sokrashhenija rabocnih mest iz-za razvitija iskusstvennogo intellekta. Sajt Forbes. [Jelektronnyj resurs]. – URL: <https://www.forbes.ru/forbeslife/481810-58-rossian-boatsa-sokrasenia-rabocnih-mest-iz-za-razvitiya-iskusstvennogo-intellekta> (data obrashhenija: 23.11.2023).
9. *Ali O. et al.* A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge*, 2023, vol. 8, no. 1, pp. 100333.
10. *Avundukluoplu P.* SAI20 2023 Gьndemi: Mavi Ekonomi ve Sorumlu Yapay Zeka. Yьl 2023, Sayь: 128, 169–176.
11. *Bhbosale S., Pujari V., Multani Z.* Advantages and disadvantages of artificial intelligence. *Aayushi International Interdisciplinary Research Journal*, 2020, vol. 77, pp. 227–230.
12. *Cann O.* Machines will do more tasks than humans by 2025 but robot revolution will still create 58 million net new jobs in next five years. *World Economic Forum*, 2018, vol. 17.
13. *Chen X.P.* The target, tasks, and implementation of artificial intelligence ethics: Six issues and the rationale behind them. *Philos. Res*, 2020, vol. 9, pp. 79–87.
14. *Chen Z.* Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 2023, vol. 10, no. 1, pp. 1–12.
15. *Gillespie N., Lockey S., Curtis C.* Trust in artificial intelligence: A five country study. – 2021.
16. *Guan H., Dong L., Zhao A.* Ethical Risk Factors and Mechanisms in Artificial Intelligence Decision Making. *Behavioral Sciences*, 2022, vol. 12, no. 9, p. 343.
17. Human-centric AI at work: A playbook for powering your organization with AI. *The Work Innovation Lab*, 2023.
18. Index I.B.M. Global AI Adoption Index IBM Corporation. 2022.
19. *Kalis B., Collier M., Fu R.* 10 promising AI applications in health care. *Harvard business review*, 2018.

20. *Kamalov F., Santandreu Calonge D., Gurrib I.* New era of artificial intelligence in education: towards a sustainable multifaceted revolution. *Sustainability*, 2023, vol. 15, no. 16, p. 12451.
21. *Marabelli M., Newell S., Handunge V.* The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *The Journal of Strategic Information Systems*, 2021, vol. 30, no. 3, p. 101683.
22. *Opletina N., Kunyaeva M.* Technology Assessment as a field of professional competency of technoscience society engineers. *2022 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2022. P. 41–44.
23. *Pham S.T.H., Sampson P.M.* The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning*, 2022, vol. 38, no. 5, pp. 1408–1421.
24. *Raporu PWC Yapay Zeka.* PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution // Bkz. <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html> (дата обращения: 22.10.2023).
25. *Russo A.* Recession and automation changes our future of work, but there are jobs coming, report says. *World Econ Forum [Online]*, 2020.
26. *Samuel J.* Response to the March 2023'Pause Giant AI experiments: an open letter'by Yoshua Bengio, signed by Stuart Russell, Elon Musk, Steve Wozniak, Yuval Noah Harari and others... *Elon Musk, Steve Wozniak, Yuval Noah Harari and others...* (March 29, 2023). 2023.
27. *Smith S.* Digital voice assistants in use to triple to 8 billion by 2023, driven by smart home devices. *Juniper Research*, 2018.
28. *Turchin A.* Structure of the global catastrophe. Risks of human extinction in the XXI century. *Lulu.com*, 2010, no. 09, pp. 67–69.
29. *Van Otterlo M.* Gatekeeping algorithms with human ethical bias: The ethics of algorithms in archives, libraries and society // arXiv preprint arXiv:1801.01705. 2018.

---

Статья поступила в редакцию 02.11.23; Принята к публикации 27.11.23.  
Авторы прочитали и одобрили окончательный вариант рукописи.

---

The article was submitted 02.11.23; accepted for publication 27.11.23.  
The authors read and approved the final version of the manuscript.

---