

Философская мысль*Правильная ссылка на статью:*

Мухаметжанова В.С., Желнина Я.Р. Этичный диалог с искусственным интеллектом: от алгоритмов до нейросетей // Философская мысль. 2025. № 10. DOI: 10.25136/2409-8728.2025.10.76178 EDN: MZXIYV URL: https://nbpublish.com/library_read_article.php?id=76178

Этичный диалог с искусственным интеллектом: от алгоритмов до нейросетей

Мухаметжанова Винера Саяровна

ORCID: 0000-0002-3164-797X

кандидат социологических наук

доцент; факультет гуманитарных и социальных наук (ФГСН); Российский университет дружбы народов им. П. Лумумбы

117198, Россия, Московская область, г. Москва, ул. Миклухо-Маклая, 10 к.2



✉ mukhametzhanova_vs@pfur.ru

Желнина Яна Романовна

независимый исследователь

117198, Россия, Московская область, г. Москва, ул. Миклухо-Маклая, 10 к.2

✉ yanajelnina@icoud.com



[Статья из рубрики "Этика"](#)

DOI:

10.25136/2409-8728.2025.10.76178

EDN:

MZXIYV

Дата направления статьи в редакцию:

07-10-2025

Дата публикации:

16-10-2025

Аннотация: Статья посвящена комплексному исследованию проблемы этичного взаимодействия с системами ИИ. В работе прослеживается эволюция феномена «умных машин»: от концептуальных представлений о формализации мышления до создания

продвинутых генеративных нейросетей. Демонстрируется прямая корреляция между усложнением технологий и выявлением широкого спектра этических вызовов. Анализируются ключевые аспекты, связанные с применением ИИ: цифровое неравенство, беспристрастность, культурное многообразие, ответственность, конфиденциальность. Авторы отмечают необходимость моральной рефлексии над проблемой ответственности за действия, совершенные с использованием высоких технологий. Особое внимание уделяется вопросам правового регулирования: рассматриваются текущие инициативы и принципы взаимодействия с искусственным интеллектом, которые уже закреплены на законодательном уровне. Авторы утверждают, что дальнейшее развитие искусственного интеллекта будет связано с переплетением этической и правовой основ, что позволит контролировать использование высоких технологий на общественном и индивидуальном уровнях. На основании теоретических обобщений и анализа нормативной базы делается вывод о необходимости симбиоза технических решений и этично-правового регулирования для создания доверительной и безопасной среды взаимодействия с системами искусственного интеллекта. В исследовании использованы исторический анализ эволюции ИИ и сравнительный метод, выявляющий связь между развитием технологий и этическими вызовами на каждом этапе генезиса ИИ. Философская рефлексия обосновывает необходимость симбиоза технических, правовых решений для безопасного использования искусственного интеллекта. Метод исторического и сравнительного анализа выявил прямую корреляцию между усложнением технологий ИИ и расширением спектра этических вызовов. Появление сложных генеративных моделей усилило проблему ответственности за действия, совершаемые с их применением. Систематизированы ключевые этико-правовые риски: от цифрового неравенства и алгоритмической предвзятости до проблем ответственности и незащищенности данных. Анализ нормативной базы показывает, что правовое регулирование не дает целостного ответа на эти вызовы. Стратегии развития ИИ демонстрируют фрагментарность и реактивный характер из-за постоянного появления новых вызовов при отсутствии их правового осмысливания. Научная новизна – в интеграционной модели, синтезирующей технические и этико-правовые стандарты. Этот подход позволит создать безопасную и доверительную среду для использования систем ИИ на общественном и индивидуальном уровнях.

Ключевые слова:

искусственный интеллект, этика искусственного интеллекта, нейросети, генезис искусственного интеллекта, цифровое неравенство, конфиденциальность данных, правовое регулирование ИИ, прозрачность алгоритмов, принцип подотчетности, ответственность

Инициативная тема № 100 939–0–000 «Применение технологий искусственного интеллекта в социологических исследованиях: практика и перспективы»

Введение

Наука – вторая природа, созданная людьми, в последние годы переживает качественный скачок в своем развитии. Высокие технологии позволили человечеству создать искусственный интеллект (далее – ИИ), имитирующий когнитивные функции человека. В настоящее время появляются все более модернизированные программы, своим поведением напоминающие человеческую мыслительную деятельность.

Искусственный интеллект – продукт выдающихся технологических достижений общества, сложная система глубоких взаимосвязей, которая широко используется в глобальных поисковых платформах и служит основой для создания интегрированных информационных структур. На сегодняшний день развитие технологий уже существенно влияет на трансформацию человеческой идентичности, преобразовывает общественные и межличностные отношения и даже затрагивает фундаментальные экзистенциональные аспекты личности.

Современное общество переходит в эпоху, когда искусственный интеллект занимает ключевую роль в процессе развития технологий. ИИ получил широкое распространение во всех областях человеческой деятельности: от медицины и образования до финансов и безопасности. В долгосрочной перспективе передовые открытия в науке смогут позволить создать более продвинутый вид искусственного интеллекта, который сможет превзойти человека не только по возможностям вычислительных мощностей, но и в других аспектах [\[16\]](#).

В настоящее время с целью автоматизации рутинных процессов, решения широкого спектра задач и прогнозирования событий широкое распространение получили нейросети – компьютерные модели, сконструированные по принципу организации нейронных сетей человека. Несмотря на высокую эффективность данных программ, они могут представлять потенциальную опасность для общества, так как их использование никак не регламентировано. Отсутствие единых стандартов порождают риски, связанные с манипуляцией данными и сбоем в работе данных систем, что может привести к серьезным последствиям как для отдельных людей, так и для всего человечества.

Возникает необходимость не только в анализе технических способностей ИИ, но и в рефлексии над отношениями человек-машина и их этическими составляющими. Важной задачей является осмысление влияния ИИ на человеческое мышление, культурные ценности и социальное взаимодействие. Среди ключевых вызовов сейчас выделяют проблемы безопасности, дискриминации, а также распределение ответственности за принимаемые ИИ решения. Данные вопросы требуют разработки комплексного подхода, объединяющего в себе как технические, так и гуманитарные знания для установления определенных правил и положений в применении технологий ИИ.

Исторические вехи становления искусственного интеллекта

Впервые термин искусственный интеллект был употреблен Джоном Маккарти в 1956 году на Дартмутской конференции. ИИ определялся как наука о создании интеллектуальных компьютерных программ [\[17, р. 2\]](#). Интеллектуальная составляющая, по мнению ученого, – это способность достигать определенной цели путем вычислительных процедур. Однако, исторически первые размышления о возможности создания мыслящих механизмов можно обнаружить еще в трудах античных мыслителей.

Концептуальное становление представлений об искусственном интеллекте напрямую связано с фундаментальными вопросами философии сознания и познания. Аристотель, рассуждая о природе мышления, заложил основы формальной логики, ставшей впоследствии краеугольным камнем развития компьютерных наук и алгоритмического мышления [\[14, р. 25\]](#). Философ первым пытался найти универсальные законы мышления, которые отразились в его учении о силлогистике. По сути, это была одна из самых ранних формальных систем дедуктивного вывода, которая представляла собой прообраз алгоритмических действий.

Средневековая философская мысль, несмотря на теоцентричность мировоззрения, также внесла значимый вклад в развитие идей о возможности механического воспроизведения мыслительных процессов. Работы Раймунда Луллия по созданию логической машины, в основе которой лежал принцип комбинаторики для выведения новых истин, можно рассматривать как первую попытку формализации процесса рассуждений и принятия решений [8, с. 61]. Он использовал вращающиеся диски с символами, означающими базовые категории, такие как Бог, мир, добродетель и т. д., и на основе их комбинаций выводил новые утверждения. Данный способ получения знаний был новаторским, и «механизм» Луллия позднее повлиял на разработку генеративных ИИ, в которых также заложен метод комбинаторики.

Эпоха Просвещения ознаменовалась появлением механистической картины мира, в рамках которой человеческое мышление начало рассматриваться как сложный, но потенциально воспроизводимый механический процесс. Труды Лейбница по универсальному исчислению заложили теоретический фундамент для будущего развития компьютерных технологий и искусственного интеллекта [14, р. 12]. Философ пытался создать формальный язык, который мог бы выражать любые идеи через комбинации символов, а также логическое исчисление для автоматизации решения споров и поиска истины, что напоминает нам современные алгоритмические процессы, используемые в машинах.

Философское осмысление природы мышления получило новый импульс в XIX веке благодаря работам Джорджа Буля, создавшего алгебру логики. Его идеи о возможности формализации мыслительных операций существенно повлияли на последующее развитие теории вычислений и компьютерных наук [14, р. 19]. Буль формализовал логику в виде алгебраической системы, используя такие операции, как И (AND), ИЛИ (OR), НЕ (NOT). Именно на основе Булевой алгебры сейчас работают все цифровые компьютеры, искусственный интеллект, а также нейросети.

Двадцатый век стал знаменательным моментом в развитии концепции искусственного интеллекта. Работы Алана Тьюринга заложили теоретические основы современной информатики и поставили фундаментальный вопрос о возможности машинного мышления. Предложенный им тест стал первым практическим критерием оценки интеллектуальности машинных систем. Его суть заключается в имитации компьютером человеческого разума, способным ввести испытуемого в заблуждение, как будто вместо компьютера с ним беседует человек [10]. Тьюринг также прогнозировал, что к 2029 машины смогут пройти данное тестирование и обмануть испытуемых.

Современный ИИ переживает революцию, движимую прогрессом в глубоком обучении, а также обработкой больших объемов информации. Ян Лекун, специалист в области ИИ, заявляет, что компьютерные разработчики вступают в новую эпоху программирования, что позволяет «не просто программировать машину для выполнения определенной задачи, а обучать ее решению более широкого круга сходных задач» [6, с. 5]. К тому же, начиная с 2020 года, широкое распространение получил генеративный ИИ: он позволяет генерировать текст, изображения, видео и голос на основе запроса пользователя. Сейчас он применяется не только в мультимедийной сфере, но и в образовании, бизнесе, науке и других областях.

Таким образом, философский анализ генезиса искусственного интеллекта позволяет выявить несколько ключевых этапов его развития. Первый этап связан с формированием теоретических предпосылок и концептуальных оснований искусственного интеллекта в

рамках классической философской традиции. Второй этап характеризуется появлением первых практических реализаций искусственных систем и формированием научного направления *artificial intelligence*. Третий этап, современный, отмечен стремительным развитием технологий машинного обучения и нейронных сетей. Мы можем предположить, что дальнейшее развитие ИИ будет тесно связано с его техническим и философским аспектами. С одной стороны, человечество стремится создать как можно больше интеллектуальных систем, применяемых в различных сферах человеческой жизнедеятельности с целью автоматизации процессов. Также весь технический потенциал будет направлен на разработку более совершенного типа ИИ, способного не только превосходить человеческий разум по объемам обработки информации, но и предоставлять такие варианты решений, которые ранее были недоступны нашему мышлению. С другой стороны, философские аспекты ИИ постоянно будут подвергаться рефлексии. Вопросы сознания, связи с человеком и положения искусственных систем в мире выйдут на первый план.

Человек и искусственный интеллект: этические дилеммы

В эпоху стремительного технологического прогресса взаимодействие человека и искусственного интеллекта становится не просто предметом научных дискуссий, но и острой этической проблемой, требующей глубокого философского осмысления. Современное общество оказалось на пороге беспрецедентных изменений, когда искусственный интеллект из теоретической концепции превратился в реальность, пронизывающую практически все сферы человеческой жизни [\[12, с. 59\]](#).

В будущем, по мере развития технологий, искусственные системы более мощного уровня могут оказаться недоступны для некоторых слоев населения [\[4, с. 50\]](#). Такая ситуация может угрожать фундаментальной идеи демократического общества – равенству. Возможно, что некоторые люди станут более совершенными: это способно привести к искусственному разделению людей и последующим крупным конфликтам. Улучшение жизни одного человека может привести к социальному неравенству и к тому, что кто-то будет иметь лучшие возможности для реализации своего потенциала. Поэтому, необходимо справедливое распределение благ и возможностей, создаваемых искусственным интеллектом. Развитие технологий не должно приводить к углублению существующего цифрового неравенства или созданию новых форм социальной дискриминации.

Фундаментальные вопросы этического характера возникают уже на стадии разработки искусственного интеллекта. Программисты и разработчики, создавая алгоритмы машинного обучения, неизбежно сталкиваются с необходимостью принятия решений, которые впоследствии будут определять поведение ИИ-систем. При этом их собственные моральные установки, культурный опыт и мировоззренческие позиции неизбежно влияют на конечный результат. Искусственный интеллект, призванный быть объективным и беспристрастным, изначально несет в себе субъективный человеческий фактор. Любая модель обучается на созданной разработчиком базе данных, которая зачастую отражает человеческие стереотипы и предрассудки [\[9\]](#). Однако, без этического надзора со стороны человека ИИ-система не способна непредвзято оценить ситуацию, так как в ее «хранилище» - базе данных набор информации всегда ограничен. Выходит, что объективность умных машин прямо коррелирует с субъективностью программиста.

Проблема культурного разнообразия в контексте глобального распространения искусственного интеллекта также заслуживает внимания. ИИ-системы могут не учитывать

культурные особенности разных стран, что создает риск культурной гомогенизации и утраты локальной идентичности. Существует опасность унификации культурных практик и утраты локальных особенностей под влиянием стандартизованных алгоритмов. По мнению российского ученого С. С. Кладько, алгоритмы искусственного интеллекта должны быть адаптируемы к различным культурным контекстам и учитывать специфику локальных этических норм и традиций [3, с. 71–72]. Следовательно, искусственный интеллект должен быть приспособлен к особенностям различных языков, социальных норм и другим духовным аспектам, чтобы обеспечить эффективное взаимодействие с пользователями из разных культурных сред.

Особое внимание вызывает проблема ответственности за действия искусственного интеллекта. Когда ИИ-система принимает решение, которое приводит к негативным последствиям, возникает закономерный вопрос: кто должен нести ответственность – разработчики, владельцы системы или сам искусственный интеллект? Технологический прогресс опережает способность общества к его осмыслению и регулированию, что создает правовой вакuum и потенциальную угрозу для общества.

Принцип ответственности разработчиков и операторов ИИ-систем требует четкого определения границ их обязательств перед обществом. В случае возникновения ошибок или сбоев в работе искусственного интеллекта должен существовать механизм установления причинно-следственных связей и определения степени ответственности каждого участника процесса создания и эксплуатации системы. Особое внимание здесь уделяется событиям, которые носят моральный характер, однако не подлежат наступлению ответственности. Например, при проявлении признаков дискриминации зачастую отсутствует виновное лицо, так как данный акт совершается ненамеренно, потому что сам ИИ не вырабатывает каких-либо намерений в процессе его применения. Распространенным подходом в таких ситуациях является распределенная ответственность, где каждый участник события несет ответственность, так как без их взаимодействия данного события бы не произошло [15]. Таким образом, в процессе использования систем технологий ИИ ответственность несут и пользователи, причем как физические, так и юридические лица, и разработчики, которые создают, тестируют или реализуют интеллектуальные системы.

Вопрос обработки персональных данных приобретает особую актуальность в контексте развития искусственного интеллекта. По мнению исследователя ИИ Н. А. Ястреб, данный аспект также затрагивает проблемы согласия на обработку данных, их защиты и конфиденциальности [13, с. 37]. ИИ-системы требуют огромных массивов данных для обучения и функционирования, что создает риски несанкционированного использования личной информации. Границы между необходимым сбором данных для улучшения работы ИИ и вторжением в частную жизнь становятся все более размытыми. Современные системы искусственного интеллекта требуют огромных массивов информации для обучения и функционирования, что создает риски несанкционированного доступа к личным данным пользователей. Необходимо разработать эффективные механизмы шифрования и анонимизации данных, обеспечивающие баланс между функциональностью ИИ и приватностью пользователей. Также подчеркивается необходимость соблюдения баланса между эффективностью работы ИИ и обеспечением конфиденциальности личной информации. Как отмечает исследователь ИИ И. И. Литвин, у искусственного интеллекта присутствует «способность обрабатывать данные без участия оператора данных, а, следовательно, без фактического распространения сведений о личной жизни субъекта персональных данных» [7, с. 116]. Однако, автор подчеркивает необходимость правового регулирования в данной сфере, так как личные

данные пользователя не защищены от передачи третьему лицу.

Исходя из вышесказанного, мы можем сделать вывод, что данные аспекты являются фундаментальными в сфере взаимодействия с ИИ. Для дальнейшего развития «умных машин» они должны быть не только включены в разработку таких систем, но и реализовываться в процессе работы агентов ИИ. Такой подход позволит предотвратить многие непредвиденные ситуации и ограничить «свободу» действий ИИ. Мы также можем предположить, что аспект ответственности на сегодняшний день стремительно выходит на первый план. С каждым днем количество преступлений, связанных с использованием технологий ИИ, растет, но санкции в контексте его применения еще не закреплены на законодательном уровне. Поэтому, необходимо регулировать использование ИИ не только в сфере внедрения этических норм в программы, но и в других областях, особенно включая правовую с целью уменьшить количество правонарушений и предостеречь пользователей от случайного или намеренного злоупотребления способностями искусственных систем.

Принципы и нормы применения искусственного интеллекта

Современное понимание искусственного интеллекта выходит далеко за рамки программного обеспечения или алгоритмических систем. ИИ представляет собой сложный комплекс технологических решений, способных к самообучению, адаптации и принятию решений на основе анализа больших массивов данных. Именно эта способность к автономному функционированию создает новые вызовы для общества и требует тщательного осмысления принципов его применения.

В Российском правовом поле законодательная основа этических принципов взаимодействия с ИИ активно разрабатывается с 2019 года. Так, в указе Президента Российской Федерации от 10.10.2019 № 490 закреплена национальная стратегия развития ИИ до 2030 года, где базовыми этическими принципами развития и использования технологий ИИ выступают: защита прав и свобод человека, безопасность, а также прозрачность. Особо важное внимание уделяется здесь первому принципу, который включает в себя как обеспечение гарантированных прав человека, включая право на труд, так и возможность получать знания для адаптации к новым условиям цифрового общества. Следовательно, этические принципы ИИ на сегодняшний день представляют не только предмет философских дискуссий, но уже являются полноценной частью национальной стратегии развития ИИ. Это означает, что данные аспекты не могут быть проигнорированы и должны быть приняты во внимание как со стороны разработчиков, так и со стороны пользователей программ.

Распоряжение правительства РФ от 19.08.2020 № 2129-р постановляет, что «развитие технологий искусственного интеллекта и робототехники должно основываться на базовых этических нормах». Эти нормы включают в себя: приоритет благополучия и безопасности человека, запрет на причинение вреда человеку, подконтрольность и соответствие закону. Таким образом, данный нормативно-правовой акт, дополняя указ Президента РФ, уже сформировал достаточно прочную законодательную основу в сфере не только государственного, но и этического регулирования искусственных систем.

В 2021 году был издан кодекс этики в сфере искусственного интеллекта, который «распространяется на отношения, связанные с этическими аспектами создания (проектирования, конструирования, пилотирования), внедрения и использования технологий ИИ на всех этапах жизненного цикла, которые в настоящее время не урегулированы законодательством Российской Федерации и/или актами технического

регулирования» [\[5, с. 4\]](#). Данный документ включает в себя все этические нормы, которые до этого лишь являлись предметом философской рефлексии и общественной дискуссии. В кодексе выделяются следующие принципы: недискриминация, безопасность, поднадзорность, ответственность, сотрудничество и достоверность информации. Такой широкий спектр позволяет достаточно полно координировать поведение ИИ и определять границы допустимого в его действиях, что формирует прочную законодательную базу для предотвращения опасного использования искусственных систем. В настоящее время, к кодексу присоединились уже более 1000 подписантов, что означает не только широкое распространение систем ИИ в разных сферах жизнедеятельности человека, но и заинтересованность отдельных лиц и компаний в соблюдении общих этических принципов при работе с искусственным интеллектом [\[5\]](#).

Отталкиваясь от содержания вышеприведенных нормативно-правовых актов, мы подробнее рассмотрим принципы применения ИИ. Одним из таких фундаментальных принципов является прозрачность алгоритмов принятия решений. Он предполагает возможность проследить логику работы системы и понять основания, на которых базируются ее выводы. Особую значимость этот аспект приобретает в контексте применения ИИ в медицине, юриспруденции и финансовом секторе, где решения системы могут иметь серьезные последствия для человека. Данный принцип предполагает обязательную этическую экспертизу алгоритмов ИИ на стадии их разработки и внедрения. Исследователь искусственного интеллекта Ю. С. Харитонова также указывает на необходимость соблюдения принципа прозрачности и объяснимости решений ИИ: «Требуется создание критериев проверки результатов работы искусственного интеллекта для целей соблюдения прав и свобод граждан, защиты государственных и общественных интересов» [\[11, с. 343\]](#).

Современные нейронные сети достигли такого уровня сложности, что даже их создатели не всегда могут объяснить процесс принятия решений системой. Это явление, получившее название «черного ящика», ставит под сомнение возможность контроля над ИИ и порождает опасения относительно его потенциальной непредсказуемости. Исследователи ИИ О. Е. Баксанский и С. Г. Сорокина обращают внимание на проблему «этической прозрачности алгоритмов ИИ», подчеркивая необходимость создания механизмов общественного контроля над развитием и внедрением систем искусственного интеллекта [\[1, с. 25\]](#). В данном контексте право человека знать, когда он взаимодействует с ИИ, и понимать принципы, лежащие в основе принимаемых системой решений, является первостепенным.

Принцип подотчетности также является одним из ключевых в контексте применения искусственного интеллекта. Подотчетность подразумевает: 1) четкое распределение ролей, ответственных лиц за разработку, внедрение и мониторинг искусственных систем; 2) документирование данных, на основе которых обучалась модель; 3) механизмы проверки – анализ алгоритмов на предмет предвзятости, ошибок и безопасности; 4) правовые последствия – штрафы и другие наказания для систем и ответственных лиц, нарушающих ответственные нормы. Принцип подотчетности необходим, чтобы решения ИИ можно было объяснить, оспорить и исправить [\[2, с. 376\]](#). Так как ошибки ИИ могут причинять людям вред, то прозрачность системы может помочь в поиске виновных, будь то разработчик программы или сама система.

Не менее важной является проблема сохранения человеческой автономии в мире цифровых технологий, где искусственный интеллект играет все более значимую роль.

Существует риск чрезмерного делегирования ответственности системам ИИ, что может привести к атрофии способности человека к самостоятельному принятию решений и моральному суждению. Если же предоставить искусственным системам полную автономность, то это может повлечь за собой серьезные проблемы: от причинения вреда как конкретным людям, так и всему обществу, до угрозы существования человеческого вида в целом. В сфере взаимодействия с ИИ необходим баланс между его самостоятельностью и человеческим контролем.

На основе анализа нормативной базы и основных принципов использования искусственного интеллекта мы можем сделать вывод, что полный перечень этических принципов уже закреплен в правовых документах, что облегчает процесс взаимодействия с ИИ и предотвращает возможные правонарушения. В обозримом будущем возможно не только издание новых законов, связанных с этикой использования искусственного интеллекта, но и установление определенных санкций ввиду злоупотребления им, что также внесет значительный вклад в процесс регулирования отношений между человеком и ИИ.

Заключение

На сегодняшний день искусственный интеллект продолжает стремительно развиваться. Такие нейросети как GPT-5, Gemini 2.0, Claude 4 и многие другие получили более мощный интерфейс, улучшили понимание и обработку текстовых запросов, уменьшили количество ошибок и стали способны к более сложным рассуждениям. Также такие модели за небольшой промежуток времени обрабатывают текст, изображение и аудио одновременно, что указывает на тенденцию к мультимодальности. Применение технологий искусственного интеллекта в науке также показывает высокую эффективность в исследованиях. Особенно это проявляется в медицине: ИИ помогают ставить точные диагнозы и ускоряют разработку новых лекарственных препаратов. Такие компании как OpenAI и DeepMind активно разрабатывают прототип AGI – сильного типа ИИ, в будущем способного выполнять более широкий круг задач. Поэтому, стремление ИИ к «всеобъемлемости» применения во всех сферах человеческой жизнедеятельности с каждым годом только продолжает возрастать.

За всю историю философских изысканий развитие концепта искусственных систем носило нелинейный, поэтапный характер. От символического подхода и первых попыток формализации мышления до качественного скачка, обусловленного доступом к большим объемам данных, искусственный интеллект прошел путь от решения узкоспециализированных задач к созданию сложных моделей, демонстрирующих признаки творческой деятельности. Ключевая закономерность, выявленная в данном исследовании, заключается в том, что по мере усложнения алгоритмов ИИ-систем происходит глубокая трансформация и масштабирование сопутствующих проблем. Если на ранних этапах развития ИИ, вызовы носили преимущественно технический характер (ограничения вычислительной мощности, недостаток данных), то современный этап связан с обострением социально-этических, правовых и философских проблем. Поэтому, настоящий успех в сфере технологий должен определяться не только скоростью машинных вычислений, но и способностью общества к своевременной и адекватной адаптации к современным реалиям.

Очевидно, что развитие искусственного интеллекта связано с рядом морально-этических проблем: обеспечение безопасности как всего общества, так и отдельного гражданина, модификация человеческого сознания и интеллекта, изменение фундаментальных мировоззренческих аспектов бытия человека. При этом человеческая мораль

переносится на роботов, включая в себя такие этические составляющие как: свобода и ответственность, равенство и справедливость, добро и зло (определение моральной оценки действий). Данные аспекты должны быть не только учтены при разработке программ, но и задействованы в процессе взаимодействия с ними, причем как со стороны искусственного интеллекта, так и со стороны пользователя. Дальнейшее развитие искусственного интеллекта будет связано как с последующей разработкой более мощных моделей ИИ, так и с формированием новой картины мира, а также с созданием новой этической реальности, где моральные нормы будут распространяться не только на людей, но и на роботов.

Формирование устойчивого правового поля в области искусственного интеллекта находится на активной стадии. Хотя уже заложены важные основы регулирования ИИ в виде законов и этических кодексов, единые стандарты на данный момент отсутствуют. Текущие инициативы направлены на решение ключевых проблем, но вместе с тем позволяют выявлять новые вызовы и способствуют развитию законодательной системы в данной сфере. В ближайшей перспективе возможно более пристальное внимание к вопросам определения правового статуса ИИ и детального регулирования в конкретных областях, где автономные системы играют значительную роль. Последующие законы должны сохранять гибкость, чтобы не тормозить технологический прогресс, и одновременно отвечать на вызовы, которые могут быть порождены все более продвинутыми моделями искусственных систем.

Искусственный интеллект — это стремительно прогрессирующая область, которая уже оказала значительное влияние на общество. Однако по мере того, как он продолжает модифицироваться, важно учитывать не только его этические и социальные последствия, но и направлять дальнейшее развитие и использование высоких технологий на благо человечества, обеспечивая, таким образом, светлое будущее самого искусственного интеллекта и реализацию его потенциала для будущих поколений.

Библиография

1. Баксанский О.Е., Сорокина С. Г. Этика искусственного интеллекта: сбалансированный подход к развитию и применению // Общество: философия, история, культура. 2025. № 1. С. 23-33. DOI: 10.24158/fik.2025.1.2 EDN: AVFHGB.
2. Бирюков Д. Н., Дудкин А. С. Объяснимость и интерпретируемость – важные аспекты безопасности решений, принимаемых интеллектуальными системами (обзорная статья) // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25. № 3. С. 373-386. DOI: 10.17586/2226-1494-2025-25-3-373-386 EDN: NHHVUJ.
3. Кладько С. С. Культурные вызовы практической этике в сфере искусственного интеллекта: проблемы и решения // International Journal of Open Information Technologies. 2024. Т. 12. № 12. С. 68-73. EDN: AIRWEF.
4. Климовицкий С. В., Осипов Г. В. Цифровое неравенство и его социальные последствия // Гуманитарные, социально-экономические и общественные науки. 2019. № 2. С. 47-51. DOI: 10.23672/SAE.2019.2.26689 EDN: ZASHNJ.
5. Кодекс этики в сфере искусственного интеллекта / принят Альянсом в сфере искусственного интеллекта 26.10.2021. Режим доступа: <https://ethics.a-ai.ru> (дата обращения: 29.06.2025).
6. Лекун Я. Как учится машина: Революция в области нейронных сетей и глубокого обучения. – М.: Альпина PRO, 2021. 423 с.
7. Литвин И.И. Особенности сбора, обработки и защиты персональных данных искусственным интеллектом // Вестник Уральского юридического института МВД России. 2021. № 4. С. 112-117. EDN: ZXXJUD.

8. Подшибякин А.М. *Ars Ex Machina: поиски искусства в постиндустриальном мире* // Вестник Нижегородского Государственного университета. 2003. № 24. С. 58-65.
9. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / гл. ред. Е. Кондукова. – 2-е изд. – СПб., 2007. 375 с.
10. Тьюринг А. М. Вычислительные машины и разум. / Пер. с англ. К. Королева. – М.: АСТ, 2018. 128 с.
11. Харитонова Ю. С. Правовые средства обеспечения принципа прозрачности искусственного интеллекта // Journal of Digital Technologies and Law. 2023. № 1 (2). С. 337-358. DOI: 10.21202/jdtl.2023.14 EDN: DXNWHV.
12. Цвыйк В.А., Цвыйк И.В. Социальные проблемы развития и применения искусственного интеллекта // Вестник Российского университета дружбы народов. Серия: Социология. 2022. Т. 22. № 1. С. 58-69. DOI: 10.22363/2313-2272-2022-22-1-58-69 EDN: YTLNJP.
13. Ястреб Н.А. Как проблема персональных данных меняет этику искусственного интеллекта? // Философские проблемы информационных технологий и киберпространства. 2020. № 1 (17). С. 29-44. DOI: 10.17726/philIT.2020.1.3 EDN: KLNWHX.
14. Davis M. *The Universal Computer: The Road from Leibniz to Turing*. – N.Y.: W.W. Norton & Company, 2000. – 239 p.
15. Floridi L. *Faultless Responsibility: on the Nature and Allocation of Moral Responsibility for Distributed Moral Actions* // Philosophical Transactions of the Royal Society. 2016. № 374 (2083). Р. 1-13.
16. Grace K., Salvatier J., Dafoe A., Zhang B. *When Will AI Exceed Human Performance? Evidence from AI Experts*. – Oxford: Oxford University Press, 2017. 21 p.
17. McCarthy J. *What is artificial intelligence?* Stanford: Stanford University, 2007. 15 p.

Результаты процедуры рецензирования статьи

Рецензия выполнена специалистами [Национального Института Научного Рецензирования](#) по заказу ООО "НБ-Медиа".

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов можно ознакомиться [здесь](#).

Предмет исследования – этичный диалог с искусственным интеллектом: от алгоритмов до нейросетей. Методологической базой исследования является метаанализ как интеграция, обобщение и философское осмысление феномена искусственного интеллекта и этических аспектов его влияние на человека и общества.

Актуальность исследования, как справедливо отмечает автор, обусловлена, тем, что в современном мире возникает необходимость не только в анализе технических способностей искусственного интеллекта, но и в рефлексии над отношениями человек-машина и их этическими составляющими.

Научная новизна исследования состоит в расширении представлений об этических аспектах взаимодействия человека и искусственного интеллекта.

Общая структура работы представлена следующими разделами: введение, результаты исследования и их обсуждение, заключение, библиография, включающая в себя 17 источников, 4 из них на английском языке. Текст статьи имеет следующую рубрикацию: «Исторические вехи становления искусственного интеллекта», «Человек и искусственный интеллект: этические дилеммы», «Принципы и нормы применения искусственного интеллекта». Содержание статьи в целом отражает ее структуру. Статья написана грамотным научным языком. Представляя и обсуждая результаты работы, автор апеллирует к трудам современных учёных, характеризующих предмет

исследования. В разделе «Исторические вехи становления искусственного интеллекта» корректно представлен философский анализ генезиса искусственного интеллекта и выявлены ключевые этапы его развития. Рассматривая этические дилеммы, порождаемые взаимодействием человека и искусственного интеллекта, автор приходит к обоснованному выводу о необходимости регулировать использование ИИ не только в сфере внедрения этических норм в программы, но и в других областях, особенно включая правовую с целью уменьшения количества правонарушений и предостережения пользователей от случайного или намеренного злоупотребления способностями искусственных систем. Анализируя принципы и нормы применения искусственного интеллекта, автор аргументировано заключает, что в обозримом будущем возможно не только издание новых законов, связанных с этикой использования искусственного интеллекта, но и установление определенных санкций ввиду злоупотребления им, что также внесет значительный вклад в процесс регулирования отношений между человеком и ИИ.

В заключении автор грамотно подводит итог работы, представляет логичные, структурированные выводы. Особенno ценным является вывод относительно того, что искусственный интеллект — это стремительно прогрессирующая область, которая уже оказала значительное влияние на общество. Однако по мере того, как он продолжает модифицироваться, важно учитывать не только его этические и социальные последствия, но и направлять дальнейшее развитие и использование высоких технологий на благо человечества, обеспечивая, таким образом, светлое будущее самого искусственного интеллекта и реализацию его потенциала для будущих поколений.

Материалы данного исследования рассчитаны на широкий круг читательской аудитории, они могут быть использованы учеными в научных целях, педагогическими работниками в образовательном процессе. Вместе с тем они представляют интерес для экспертов и аналитиков.

В качестве недостатков работы следует отметить отсутствие во введении характеристики степени научной разработанности проблемы и методологии исследования.