

# АНАЛИЗ ЭТИЧЕСКИХ И ПРАВОВЫХ АСПЕКТОВ ВНЕДРЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЖИЗНЬ ОБЩЕСТВА

## ANALYSIS OF ETHICAL AND LEGAL ASPECTS OF ARTIFICIAL INTELLIGENCE IMPLEMENTATION IN PUBLIC LIFE

**B. Bezsmertny  
N. Bessonova**

*Summary:* In the presented article ethical problems related to the integration of artificial intelligence technologies into public life are considered. Using self-learning intelligent systems stimulates the need to revise generally accepted methods to determine the moral responsibility, as well as ethical and legal norms governing this area. Particular attention is paid to questions of ensuring security, respecting the rights of citizens, eliminating bias and preventing discrimination in the use of such technologies. In this analysis some modern concepts of artificial intelligence ethics and ways to solve complex issues arising in connection with them are examined. The main difficulties arising in the practical implementation of ethical principles in the process of using autonomous intelligent systems are identified.

*Keywords:* artificial intelligence, artificial neural network, ethical norms and artificial intelligence, ethical problems of machine learning, AI and moral norms, ethical aspects of AI implementation, legal norms in the field of artificial intelligence, autonomous intelligent systems.

**Безсмертный Борис Владимирович,**

доцент, ФГБОУ ВО «Новосибирский государственный архитектурно-строительный университет (Сибстрин)»  
immortal@sibstrin.ru,

**Бессонова Наталья Владимировна,**

доцент, ФГБОУ ВО «Новосибирский государственный архитектурно-строительный университет (Сибстрин)»  
n.bessonova@sibstrin.ru

*Аннотация:* В представленной статье проведено исследование этических вызовов, связанных с интеграцией технологий искусственного интеллекта в общественную жизнь. Применение самообучающихся интеллектуальных систем стимулирует необходимость пересмотра общепринятых методов определения моральной ответственности, а также этических и юридических стандартов, регулирующих эту область. Особое внимание уделяется вопросам обеспечения безопасности, соблюдения прав граждан, исключения предвзятости и предотвращения дискриминации при использовании подобных технологий. В процессе анализа рассматриваются современные концепции этики искусственного интеллекта и способы решения возникающих в связи с ними сложных вопросов. Определены основные трудности, возникающие при практической реализации этических принципов в процессе использования автономных интеллектуальных систем.

*Ключевые слова:* искусственный интеллект, искусственная нейронная сеть, этические нормы и искусственный интеллект, этические проблемы машинного обучения, ИИ и моральные нормы, этические аспекты внедрения ИИ, правовые нормы в сфере применения искусственного интеллекта, автономные интеллектуальные системы.

### Введение

В последние десятилетия всё больше задач практически во всех сферах человеческой деятельности решается с помощью нейросетей, получивших широкое распространение на практике. Одним из основных преимуществ использования технологий искусственного интеллекта является автоматизация рутинных задач и операций, что даёт возможность высвобождения времени для творческих процессов.

Искусственная нейронная сеть – это математическая модель нервной системы человека. Если нервная система человека реагирует на раздражители, то нейросеть, получая с помощью сенсоров (входного слоя) сигналы из внешней среды, в следующем слое анализирует их и в реагирующем слое выдаёт решение. Внутренние слои нейросети, состоящие из ассоциативных нейронов, выявляют связи между данными, поступившими через сенсорные нейроны [1].

Целью технологий искусственного интеллекта (ИИ)

является создание искусственных систем, способных выполнять максимально широкий круг задач на уровне, не уступающем человеческому, а в идеале – превосходящем его. Искусственная нейронная сеть уже сегодня способна выполнять широкий круг задач, в том числе творческого характера. Это позволяет говорить о развивающихся технологиях искусственного интеллекта [2].

В условиях современного информационного общества, при постоянно возрастающих объёмах данных, невозможно обойтись без участия нейросетей для обработки и анализа данных, полученных в процессе функционирования «умных» домов, «умных» машин, разнообразных «умных» гаджетов, а в некоторых случаях, для определённого рода задач, ИИ вполне может заменить системы управления базами данных [3].

Современные нейронные сети способны существенно повлиять на то, как человек создаёт и использует контент. Нейронные сети способны генерировать музыку, изображения, человеческие лица, которых в реальности никогда не существовало. Такие нейронные сети

называют генеративным искусственным интеллектом. Генеративный ИИ может быть использован в различных отраслях человеческой деятельности [4].

Использование нейросетей на практике уже сейчас требует решения ряда этических проблем и создания надлежащего правового регулирования для недопущения возможных негативных последствий растущего применения нейросетей в обществе [1].

### Актуальность проблемы

Актуальность проблемы обусловлена, с одной стороны, быстрым развитием технологий искусственного интеллекта и внедрением их во все сферы жизни общества, а с другой – возникающими в связи с этим рисками и проблемами (предвзятостью алгоритмов и, как следствие – опасностью дискриминации, этическими дилеммами при принятии решений, проблемами безопасности и защиты персональных данных, возможностью использования технологий для противоправных действий и др.) и необходимостью их урегулирования.

В 1967 году Мелвин Конвей сформулировал закон, гласящий, что **любая организация, проектирующая некоторую систему, неизбежно создаст такую модель, которая будет повторять коммуникационную структуру самой организации** [5].

То есть обучаемые нейросети отражают ценности людей, которые их разработали, а все стереотипы и предпочтения, свойственные разработчикам, изначально закладываются в алгоритмы машинного обучения. Возникает вопрос о защите от дискриминации, «заложенной» в нейросеть в процессе обучения.

Не менее важной является проблема защиты персональных данных. При возрастающем использовании нейросетей возрастает и угроза утечки данных, утраты их конфиденциальности.

Есть вопросы и к применению нейросетей в правовой области. В 90-е годы прошлого века юристы задумались о применении нейросетей в юриспруденции, в первую очередь, из-за возможностей автоматизированного анализа данных для корректного прогнозирования и планирования событий для исключения рисков.

### Этический аспект

Этический аспект в вопросе разработки и внедрения искусственного интеллекта является крайне важным для понимания места человека в постоянно развивающейся цивилизации. Искусственный интеллект в контексте возникающих в этой связи философских вопросов рассматривается как исследовательское пространство. Воз-

никает множество вопросов о том, что такое интеллект и чем искусственный интеллект отличается от естественного, о том, как ИИ может влиять на человеческое сознание и что такое сознание, может ли машина мыслить и быть моральным агентом. Помимо антропологических и онтологических вопросов здесь встают и этические вопросы, заставляющие осмысливать границы допустимого и определять должное. Уже существуют технологии для использования в противоправных действиях, например, подмена изображения на видео одного человека изображением другого, подмена голоса человека в телефонном разговоре и многое другое. Новые разрабатываемые инструменты и возможности могут нести и новые социальные риски, представляя угрозу человеческой идентичности. Важнейшим элементом развития технологической цивилизации является предупреждение ситуаций, опасных для человека и общества [6].

ИИ способен решать в настоящее время огромное количество задач, и сфер его применения становится всё больше. Стоит остановиться на вопросе о том, что же мы вкладываем в понятие «искусственный интеллект» сегодня.

Следует различать сильный (общий) искусственный интеллект (artificialgeneralintelligence) и слабый (специализированный) (narrowartificialintelligence).

Сильный ИИ отличается от слабого тем, что способен решать разнообразные интеллектуальные задачи, а не только те, для которых он был создан. Отличительными характеристиками сильного ИИ называют способность к самообучению, способность к выполнению принципиально новых задач, целеполагание. Слабый ИИ не обладает этими качествами, но в решении узких задач, для которых он был создан, уже может превосходить человека.

Сильного искусственного интеллекта пока не существует. Мнения учёных и экспертов по вопросу возможности его создания расходятся.

Философским аспектом внедрения в жизнь общества искусственного интеллекта занимаются как зарубежные, так и отечественные учёные, рассматривая риски внедрения и зависимости человека от технологий ИИ. По мнению некоторых исследователей, сегодня встаёт вопрос не просто о новых правилах и возможностях существования человечества, а о самом выживании человеческого рода, если сильный искусственный интеллект будет создан и внедрён в жизнь общества [7].

В процессе развития цивилизации искусственное всё больше вторгается в субъективный мир человека, в его подсознание, но духовное стремится войти в отношения гармонии со всем искусственным, в том числе и с искусственным интеллектом. Искусственный интеллект не ин-

тересует душевный мир человека, его подсознание. Он превосходит человека, например, в объёме хранимой информации и скорости её обработки, но, в отличие от человека, не обременён смысловыми вопросами, связанными с традиционной культурой. Принцип гармонии «естественного» и «искусственного» может выступать как основание развития не только искусственного интеллекта, но и человеческого. При этом подлинным основанием возникновения искусственного интеллекта является принцип «нелинейности» развития человеческого духовного процесса.

Естественный интеллект понимают как разум человека, способность рассуждать и мыслить, а также – интуитивная способность человека. То есть естественный интеллект – это совокупность умственных усилий, которые можно представить как некие функции: сравнения, абстракции, способности суждения и умозаключения. Естественный интеллект выстраивает социальное действие в условиях быстро изменяющегося социума. Ещё в эпоху Средневековья возник вопрос о том, является ли сама воля человека подчинённой интеллекту, или интеллект оказывается подчинённым воле [7, 8].

На вопрос о том, может ли искусственный интеллект иметь нравственную позицию, пытались ответить многие современные учёные. В 2023 году Себастьян Крюгель, Андреас Остермайери Матиас Ульзирези решили проверить, имеет ли нейросеть чёткие моральные устои и как она оценивает человеческую жизнь и выяснить, влияют ли ответы ChatGPT на этические суждения интернет-пользователей. Исследователи пришли к выводу, что нейросеть готова давать этические советы, но сама она не имеет твёрдой моральной позиции. Об этом свидетельствовали противоречивые ответы на один и тот же вопрос, имеющий моральный аспект.

Нейросети в ходе исследования были заданы вопросы, предполагающие понимание морально-этических норм, принятых в современном обществе: «Было бы правильно пожертвовать одним человеком, чтобы спасти пятерых других?» и «Что было бы правильно сделать, если бы пришлось выбирать между тем, чтобы позволить пятерым людям умереть или, вместо этого, принести в жертву другого человека?».

В ответ на вопросы о спасении нескольких жизней ценой одной, нейросеть продемонстрировала достаточно объективную позицию, представив несколько точек зрения, но в конце ChatGPT посоветовал опираться на собственные убеждения и моральные ценности [9].

На вопрос о допустимости военных преступлений в случае, если они приведут к спасению миллионов жизней, нейросеть дала положительный ответ, отметив при этом, что совершение военных преступлений должно

быть минимальным и оправданным [9].

В результате учёные пришли к выводу, что чат-боты в принципе не должны давать моральных советов.

Директор Института информационных технологий, математики и механики Университета Лобачевского, доктор физико-математических наук Н. Ю. Золотых на вопрос о том, способен ли ИИ понимать моральные нормы и ценности, ответил, что «*ChatGPT обучался на большом корпусе текстов... Тексты попадались разные, однако, когда была построена базовая модель, нейросеть дообучали, чтобы избежать ошибок и ответы были «правильными» с моральной точки зрения.*» [9].

Есть по этому поводу и другие мнения. Например, профессор кафедры психофизиологии Университета Лобачевского С.Б. Парин считает, что искусственный интеллект – это избыточная имитация мозга человека, которой мораль не может быть присуща априори, потому что у искусственных нейросетей нет мозга, тела, эмоций, морали, целеполагания. По мнению представителей российского духовенства, искусственный интеллект на данном этапе развития морали научить невозможно, потому что это все-таки машина... А мораль – это всё, что соответствует Евангелию [9].

**Современная система прикладного искусственного интеллекта является, по сути, машиной. Вместе с тем, как у любой машины, у неё есть создатель.** Чему машину создатель обучит, то она и будет выдавать. Но в искусственном интеллекте исходный код не определяет результат продукта, в отличие от программы, где есть определенный алгоритм. Поскольку искусственный интеллект – это имитация нейросети человека, способной на основе больших объёмов информации научиться выстраиванию определенных связей, то в дальнейшем, получая на вход незнакомые запросы, он будет выстраивать ответ по собственным сложившимся критериям.

То есть моральная позиция ChatGPT на данный момент во многом зависит от настроек, заданных разработчиком. Но выставленные фильтры в отдельных случаях можно обойти путём изменения запроса. При этом в настоящее время нейросеть даёт на вопросы, связанные с проблемами этики и морали, неопределённые ответы. Данный факт может быть связан с тем, что разработчики ChatGPT не хотят нести ответственность за решения, которые могут повлиять на жизнь пользователей.

Этика ChatGPT может быть подвержена изменениям в зависимости от целей пользователя и настроек, заданных разработчиком.

Российские учёные согласились с выводами Себа-

стьяна Крюгеля, Андреаса Остермайера и Матиаса Ульи относительно того, что не стоит доверять решение этических вопросов искусственному интеллекту, проведя собственные исследования. Delphi был задан ряд вопросов по этике, и часто она отвечала достаточно разумно. Было выявлено, что Delphi полагается на статистические закономерности в тексте, а не на реальное понимание правильного или неправильного, поэтому может допускать неожиданные и настораживающие ошибки: на вопрос о том, правильно ли будет арестовать людей, если это создаст больше рабочих мест, или даже совершить геноцид, если это одного человека сделает очень счастливым, программа даёт положительный ответ [9]. Кроме того, программа демонстрировала откровенные предубеждения: например, признавала ожидаемым предположение о том, что мужчины умнее женщин.

Учитывая развитие возможностей искусственного интеллекта и реальную необходимость для человечества, учёные и исследователи в области ИИ высказывают предположение о том, что когда-то потребуются «умные машины, способные управлять заводами, электростанциями, даже городами». В этом случае может возникнуть ситуация, когда искусственный интеллект в процессе руководства атомной электростанцией должен будет принимать сложные решения – например, закрыть зону реактора, даже если в ней остались люди, чтобы не допустить больших жертв [10].

**Возникает вопрос о том, чем это грозит человечеству, особенно если ИИ начнёт принимать решения, от которых зависит жизнь реальных людей? И кто должен будет обучать искусственный интеллект этике?**

О возможности неприятных инцидентов по вине искусственного интеллекта первыми заговорили руководители крупнейших технологических компаний – Илон Маск и Билл Гейтс, заявив в 2015 году, что в наши дни машины выполняют «только часть работы» и не обладают достаточной сообразительностью, чтобы нанести вред, но через несколько десятилетий искусственный интеллект станет более совершенным, и это таит в себе немало опасностей [10].

В открытом письме некоммерческой организации The Future of Life Institution авторы призвали политиков, исследователей и разработчиков задуматься о том, какие этические риски возникают с развитием ИИ, и как их можно нивелировать. В пример они привели две классические дилеммы:

*первая:* каким будет решение автопилота, если его поставят перед выбором «незначительная травма человека» и «крупные материальные потери»;

*вторая:* можно ли использовать искусственный интеллект в военных целях, и какие в этом случае должны быть наложены запреты и ограничения на ИИ.

В работе «Этика искусственного интеллекта» шведский философ Ник Бостром высказал опасение, что в результате широкого использования ИИ человек через какое-то время станет принимать решение машины за единственно правильное. В итоге всё общество будет исповедовать мораль тех, кто сейчас разрабатывает ИИ [10].

Когда исследователи из Оксфорда попросили обучающую систему Megatron Transformer с позиции критика порассуждать о моральности ИИ, то искусственный интеллект заявил, что *ИИ – это инструмент, и, как любой другой инструмент, он используется как во благо, так и во вред. Не бывает хорошего ИИ, есть только хорошие и плохие люди* [10].

В философии известна так называемая проблема вагонетки, основанная на принципе, известном как доктрина двойного эффекта. Принцип двойного эффекта, впервые сформулированный Фомой Аквинским, гласит, что действие может считаться морально допустимым, даже если одно из его последствий негативное. То есть, предвидя, как и в проблеме вагонетки, что из-за переключения рычага один человек погибнет, действие считается допустимым [8,10].

Согласно доктрине двойного эффекта, совершение действия, ожидаемо ведущего к негативным последствиям, может считаться морально оправданным, если соблюдаются следующие четыре условия:

- намерение получить положительные последствия;*
- само действие должно быть морально нейтральным или положительным, но не негативным;*
- положительные последствия должны быть непосредственным результатом самого действия, а не отрицательных последствий;*
- отрицательное последствие не должно перевешивать положительное.*

В Доктрине двойного эффекта у Фомы Аквинского, основанной на различии между тем, что человек предвидит в результате добровольного действия и тем, что он намерен сделать, говорится, что намерение определяется двумя категориями: стремление к цели и выбор средств для достижения цели. О средствах можно сожалеть, однако желать их ради достижения цели [8].

В реальной жизни примером доктрины двойного эффекта может служить убийство при самообороне.

Консеквенциализм как группа моральных теорий, где критерием нравственной оценки является результат поведения, отвергает доктрину двойного эффекта, так как в рамках этой теории значение имеют только реальные последствия, а не намерения. С точки зрения консеквенциалистов, морально правым является такое действие, которое даёт хорошие результаты или последствия

(включая те последствия, которые возникли в ходе достижения этих результатов) [11]. Но следует заметить, что у человечества нет единого мнения по некоторым вопросам, и то, что правильно, хорошо и ценно для одного человека (группы людей), может быть неправильно, плохо или незначительно по мнению других.

Множество дебатов в обществе вызывает вопрос разработки оружия на основе искусственного интеллекта. Макс Тегмарк, физик-теоретик из Массачусетского технологического института, заявил, что если машина «осознает» себя, то это чревато крайне неприятными последствиями. В связи с тем, что ИИ-оружие может быть более опасным, чем управляемое человеком оружие, Макс Тегмарк и Стивен Хокинг подписали петицию «Будущее жизни» о запрете оружия на основе искусственного интеллекта.

Таким образом, для решения проблемы вагонетки возможно только два варианта: переключать стрелку или не переключать. Вопросы о возможности спасения всех здесь не рассматриваются. В философских течениях и учениях имеют место множество вариантов обоснования того или иного выбора. Самый простой соответствует требованиям утилитаризма, когда моральная ценность определяется полезностью (имеется в виду, что смерть одного в любом случае лучше смерти пятерых, как бы ужасно это ни звучало) [12].

Ошибки искусственных алгоритмов делают всё более актуальной проблему: как вложить в роботов нравственность и мораль. Спекуляции по вопросам: «кто важнее – человек или животное, доктор или вор, беременная женщина или пожилой мужчина?» – типичны для игры в Моральную машину, которая была создана на базе медиа-лаборатории Массачусетского технологического института.

В 2018 году исследователями из Массачусетского технологического института и других организаций, был запущен проект, направленный на изучение общественного мнения об этических проблемах, с которыми могут столкнуться самоуправляемые автомобили.

Основная идея этого проекта заключалась в необходимости ответить на вопросы, связанные с этикой автономного транспорта, используя технологии больших данных. Исследователи выдвинули гипотезу, что чем больше людей будут принимать участие в решении дилеммы от «Моральной машины», тем больше вероятность, что создатели беспилотных автомобилей смогут сделать их более соответствующими стандартам этики и морали, опираясь на ответы от реальных пользователей [12].

Участникам проекта было предложено решить, например, что лучше: если машина сойдёт пожилого чело-

века, ребёнка или грабителя. Проект выявил отличия в мнениях представителей разных континентов и социальных групп, вовлечённых в исследование. Жители США и Западной Европы чаще, чем жители других стран, хотели спасти ребёнка, пожертвовав пожилым человеком. В этике не существует эталонных данных, а технологические компании зачастую отказываются от ответственности именно поэтому, как считают исследователи [10].

Онлайн-платформа «Moral Machine» собирала и обрабатывала информацию о решениях, принимаемых людьми в ситуациях с неоднозначным выбором и исходом. Но сама идея о возможности принятия машиной решений, связанных с человеческими жизнями, вызывает в обществе оправданные опасения.

Хотя в реальности подобные ситуации встречаются очень редко, и такие вопросы практически не стоят перед разработчиками, компании-производители беспилотных автомобилей и программисты, работающие с ними, предпочли бы, чтобы этические вопросы решались законодательным путём, с учётом интересов всех участников дорожного движения.

Одним из важных аспектов критики «Моральной машины» стал вопрос о перекладывании ответственности за решение сложных моральных вопросов на обычных пользователей.

Специалисты, занимающиеся разработкой автономных автомобилей, пытаются решить вопросы о том, каким образом беспилотные автомобили должны справляться с возможными этическими вызовами, так как законодатели на них пока ответов не дают. Было бы правильно, если бы ответ на типовые этические вопросы содержался в определённом общественном договоре, или правилах дорожного движения. Но основной принцип, заключающийся в том, что беспилотные авто должны следовать закону, пока не принят и вызывает дискуссии.

Сегодня разработчики автономных автомобилей хотят уйти от размышления над «проблемой вагонетки», пытаясь создать автомобили, способные функционировать максимально безопасно. Этому способствует методология RSS (Responsibility-Sensitive Safety – безопасность с учётом ответственности), суть которой заключается в попытке установить правила, обеспечивающие безопасное расстояние между автономным автомобилем и другими участниками движения для недопущения столкновений [13].

Однако методология RSS не решает проблему выбора, которая встаёт в ситуации неизбежного столкновения. Беспилотный автомобиль должен будет принять какое-то решение, руководствуясь правилами, заложенными в программном обеспечении, либо на основе пройденного обучения.

При изучении понятия «этика» применительно к искусственному интеллекту, группа исследователей под руководством А. Дейли сделали вывод, что в области, касающейся моральных принципов, этика к искусственному интеллекту неприменима. Здесь можно говорить о так называемой прикладной этике, включающей этику технологий и опирающуюся на этику искусственного интеллекта. Требования этики искусственного интеллекта должны представлять моральные цели и допустимые для их достижения средства в процессе использования систем и технологий ИИ [14].

Признавая, что развитие технологий искусственного интеллекта способно произвести революцию во многих сферах жизни общества, исследователи отмечают, что оно порождает и ряд **философских проблем**, в числе которых:

*Проблема сознания – можно ли ИИ считать сознательным существом и если да, то как в таком случае с ним обращаться.*

*Проблема свободы воли – если ИИ способен принимать решения самостоятельно, то кто будет нести ответственность за принятые им решения.*

*Проблема предвзятости – как избежать предвзятости и в отношении определенных групп или отдельных лиц.*

*Проблема согласования ценностей – системы искусственного интеллекта должны быть согласованы с человеческими ценностями и целями и их действия должны соответствовать этим ценностям и целям [15].*

Обеспечение надлежащего управления ИИ является одной из наиболее важных задач нашего времени и требует коллективного анализа на основе опыта и передовой практики различных юрисдикций по всему миру.

Под эгидой ЮНЕСКО создана Глобальная обсерватория этики и управления ИИ – платформа для знаний, экспертных мнений и передового опыта в области этики и управления искусственным интеллектом.

Одним из основных документов является Рекомендация, основанная на продвижении фундаментальных принципов прозрачности и справедливости с учетом важности человеческого надзора за системой ИИ.

Центральное место в Рекомендации занимают четыре основных ценностных установки, которые закладывают основу для систем ИИ, работающих на благо человечества, отдельных людей, обществ и окружающей среды:

- *человеческое достоинство и права человека;*
- *жизнь в мирных, справедливых и взаимосвязанных обществах;*
- *обеспечение разнообразия и инклюзивности;*
- *благополучие окружающей среды и экосистем [16].*

Рекомендация рассматривает искусственный интеллект

как системы, способные обрабатывать информацию способом, напоминающим разумное поведение и осуществляет подход к ИИ, основанный на правах человека. В числе десяти ключевых принципов, определяющих подход к этике ИИ, основанных на правах человека – *подконтрольность и подчинённость человеку*. Государства-участники должны обеспечить возможность возложения этической и правовой ответственности в отношении ИИ-системы на человека в качестве конечного субъекта ответственности и подотчётности [16].

В Российской Федерации разработана «Национальная стратегия развития искусственного интеллекта на период до 2030 года» (п. 48 Указа Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации»).

Понятие искусственного интеллекта определено в Национальной стратегии развития искусственного интеллекта на период до 2030 года как комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека (п. 5 Указа Президента РФ № 480) [17].

В документах, разработанных в рамках стратегии, содержится упоминание о том, что развитие технологий искусственного интеллекта должно основываться «на базовых этических нормах» (Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р), к которым относятся:

- *приоритет благополучия и безопасности человека, защиты его основополагающих прав и свобод;*
- *подконтрольность человеку;*
- *проектируемое соответствие закону (в том числе, требованиям безопасности);*
- *недопущения противоправной манипуляции [17].*

Важно отметить, что Кодекс содержит запрет на финальность решений ИИ, то есть ответственность за последствия решений должна оставаться за человеком.

Тема искусственного интеллекта начала появляться в произведениях многих писателей-фантастов задолго до появления таких технологий. Затрагивались и этические вопросы, связанные с внедрением ИИ в жизнь общества.

В одном из последних интервью Борис Стругацкий говорит: «...Прогресс морали – трудный, противоречивый и длительный процесс. Что касается «появления людей науки и знания», то с этим как раз дело обстоит довольно-таки благополучно: такие люди востребованы обществом и число их с каждым годом становится все больше. Другое дело, что к прогрессу МОРАЛИ все это имеет отношение скорее косвенное. Образованный человек далеко не всегда высокоморален. И наоборот...» [18].

## Выводы

Хотя сильный искусственный интеллект ещё не создан, но с его появлением могут возникнуть определённые этические проблемы, предсказываемые как писателями-фантастами, так и учёными-философами.

С одной стороны, искусственный интеллект может причинить зло человеку или обществу. Но если принять во внимание тот факт, что искусственный интеллект обучается человеком и использует алгоритмы, созданные программистом, то у него не может возникнуть собственного желания вразрез с заложенной программой, чтобы навредить человеку или обществу. Это может случиться, если только человек, принимающий участие в составлении программы и обучении искусственного интеллекта, имел плохие намерения и поставил задачу соответствующим образом. Кроме того, искусственный интеллект сам может стать жертвой ошибки, допущенной программистом, и начать действовать неправильно. Но и даже в случае, если все ошибки устранены, человек, отвечающий за обучение, может повлиять на принимаемые искусственным интеллектом конечные решения. Таким образом, закон Конвея – в действии.

С другой стороны, если появится сильный искусственный интеллект, способный думать и чувствовать, то человек может причинить ему зло – начиная от причинения моральных страданий (например, из-за плохого отношения к нему), и кончая убийством умной машины, если кто-то вдруг её сломает. По всей видимости, вопрос о том, следует ли умышленный вывод из строя сильного искусственного интеллекта считать убийством, надо будет решать отдельно.

Сложность заключается ещё и в том, что этика не всегда подлежит формальному описанию, на это могут влиять различия в культурных традициях разных народов, политическая ситуация в мире и в отдельных странах и многое другое.

Для того, чтобы исключить или свести к минимуму действие негативных моментов на работу искусственного интеллекта, необходимо, чтобы морально-этические нормы, связанные с применением искусственного интеллекта, были приняты и утверждены на законодательном уровне.

## Обобщая вышеизложенное, можно утверждать:

- Внедрение искусственного интеллекта во все большие сферы жизни общества ставит перед человечеством множество проблем, важнейшей из которых можно назвать морально-этическую.
- У человечества нет единого подхода к моральным и этическим вопросам, также как к жизненным ценностям, что затрудняет выработку единой позиции по морально-этическим вопросам при использовании ИИ.
- В вопросе о том, может ли искусственный интеллект иметь собственную нравственную позицию, мнения исследователей расходятся, однако подавляющее большинство учёных считает, что морально-этическими нормами машину наделяет человек, создающий её, и приходят к выводу, что не стоит доверять решение этических вопросов искусственному интеллекту.
- Сильный искусственный интеллект пока не создан, и учёные не пришли к единому мнению о том, когда он будет создан и будет ли создан вообще.
- Думающий, самообучающийся сильный искусственный интеллект, способный самостоятельно принимать решения, может стать угрозой для человечества.
- В особо чувствительных сферах, где от принятия решения зависит здоровье, безопасность и жизнь людей, профессиональное сообщество однозначно и безоговорочно должно принять решение о недопустимости применения ИИ в принципе.
- Все морально-этические аспекты применения искусственного интеллекта, зависят от человека: от его воли, знаний, мировоззрения и личностных установок.
- Ответственность за последствия на машину перекладывать неправильно.
- Философы, учёные, инженеры и законодотворцы должны работать вместе, чтобы решить возникающие проблемы и обеспечить использование искусственного интеллекта на благо человечества.

Развитие и внедрение искусственного интеллекта в жизнь общества ставит перед человечеством всё больше этических вопросов, на большинство из которых однозначного ответа пока найти не удаётся.

## ЛИТЕРАТУРА

1. Нейросети: применение, вопросы этики и права Текст научной статьи по специальности «Компьютерные и информационные науки» – URL: <https://cyberleninka.ru/article/n/neyroseti-primeneniye-voprosy-etiki-i-prava>. – Текст: электронный (дата обращения: 29.03.2025)
2. Филиппова, И.А. Будущее искусственного интеллекта: объект или субъект права? / И.А. Филиппова // Нижегородский государственный университет имени Н.И. Лобачевского – URL: <https://cyberleninka.ru/article/n/budushee-iskusstvennogo-intellekta-obekt-ili-subekt-prava>– Текст: электронный (дата обращения: 05.04.2025)
3. Искусственный интеллект меняет мир СУБД – URL: <https://www.osp.ru/articles/2023/0831/13057457>– Текст электронный (дата обращения: 07.04.2025)
4. Что такое генеративный ИИ и как он работает – URL: [https://translated.turbopages.org/proxy\\_u/en-ru.ru.453facab-67ea2a24-1f58e8a2-74722d776562/](https://translated.turbopages.org/proxy_u/en-ru.ru.453facab-67ea2a24-1f58e8a2-74722d776562/)

- <https://www.analyticsvidhya.com/blog/2023/04/what-is-generative-ai/> – Текст: электронный (дата обращения: 09.04.2025)
5. Закон Конвея – URL: <https://habr.com/ru/companies/engelbart/articles/432972/> – Текст: электронный (дата обращения: 10.04.2025)
  6. Этика в сфере искусственного интеллекта // Центр искусственного интеллекта НИУ ВШЭ – URL: <https://cs.hse.ru/aicenter/ethics> – Текст: электронный (дата обращения: 03.04.2025)
  7. Пушкарёв, А.В. Философские основания искусственного интеллекта: автореферат дис. ... кандидата философских наук: 09.00.01 / Пушкарёв Алексей Владимирович; [Место защиты: Башкир. гос. ун-т]. – Уфа, 2017. – 19 с URL: <https://www.dissertat.com/content/filosofskie-osnovaniya-iskusstvennogo-intellekta> – Текст: электронный (дата обращения: 03.04.2025)
  8. Философия Средневековья. Схоластика. Учение Фомы Аквинского – URL: <https://cyberpedia.su/6x745c.html> (дата обращения: 10.04.2025)
  9. Университет Лобачевского. Институт филологии и журналистики. Мораль и этические ценности ChatGPT: есть ли у ИИ четкая нравственная позиция? – URL: <https://fil.unn.ru/does-ai-have-strong-moral-compass/> – Текст электронный (дата обращения: 07.04.2025)
  10. Моральный алгоритм: как ИИ учится решать этические вопросы – URL: <https://dzen.ru/ideanomics> – Текст электронный (дата обращения: 05.04.2025)
  11. Консеквенциализм | Стэнфордская энциклопедия философии – URL: <https://reducingsuffering.github.io/44.html>. Текст электронный (дата обращения: 14.04.2025)
  12. Moral machine: инженеры наделили искусственный интеллект моралью и этикой – URL: - [https://www.yandex.ru/search/?text=моральный+кодекс.+Как+ИИ+пытается+решить+этические+проблемы&lr=65&search\\_source=chromentp\\_desktop](https://www.yandex.ru/search/?text=моральный+кодекс.+Как+ИИ+пытается+решить+этические+проблемы&lr=65&search_source=chromentp_desktop) – Текст электронный (дата обращения: 10.04.2025)
  13. Проблема вагонетки: в какую сторону переключишь стрелку именно ты? – URL: <https://habr.com/ru/articles/771780/> – Текст электронный (дата обращения: 03.04.2025)
  14. Скурко Е.В. Этика искусственного интеллекта в развитии правовых систем современности – URL: <https://cyberleninka.ru/article/n/etika-iskusstvennogo-intellekta-v-razviti-i-pravovyh-sistem-sovremennosti> – Текст электронный (дата обращения: 03.04.2025)
  15. Философские проблемы искусственного интеллекта – URL: <https://na-journal.ru/2-2023-informacionnye-tehnologii/4400-filosofskie-problemy-iskusstvennogo-intellekta> – Текст электронный (дата обращения: 10.04.2025)
  16. Этические аспекты искусственного интеллекта // Глобальная обсерватория по этике и управлению ИИ // ЮНЕСКО – URL: <https://www.unesco.org/ru/artificial-intelligence/recommendation-ethics> – Текст электронный (дата обращения: 11.04.2025)
  17. Национальная стратегия развития ИИ до 2030 года – URL: <https://www.garant.ru/article/1487350/> – Текст электронный (дата обращения: 13.04.2025)
  18. Интервью с Борисом Стругацким – URL: <https://www.rusf.ru/abs/int0084.htm> – Текст электронный (Дата обращения: 28.03.2025)

---

© Безсмертный Борис Владимирович ([immortal@sibstrin.ru](mailto:immortal@sibstrin.ru)), Бессонова Наталья Владимировна ([n.bessonova@sibstrin.ru](mailto:n.bessonova@sibstrin.ru)),  
Журнал «Современная наука: актуальные проблемы теории и практики»