

Философская мысль

Правильная ссылка на статью:

Авдюнин В.А. Этические пробелы в этике искусственного интеллекта // Философская мысль. 2025. № 12. DOI: 10.25136/2409-8728.2025.12.77033 EDN: FICGLO URL: https://nbpublish.com/library_read_article.php?id=77033

Этические пробелы в этике искусственного интеллекта

Авдюнин Василий Андреевич

ORCID: 0000-0001-5220-7255

старший преподаватель; кафедра философии; ФГБОУ ВО Вологодский государственный университет
аспирант; кафедра философии; ФГБОУ ВО Вологодский государственный университет

160033, Россия, Вологодская обл., г. Вологда, ул. Космонавта Беляева, д. 20, кв. 19

✉ vasya.avdiunin@yandex.ru



[Статья из рубрики "Информация и идеи"](#)

DOI:

10.25136/2409-8728.2025.12.77033

EDN:

FICGLO

Дата направления статьи в редакцию:

26-11-2025

Дата публикации:

03-12-2025

Аннотация: Развитие нейросетей и их активное внедрение в различные сферы экономики поставило задачу этической оценки и правового регулирования разработки и применения систем искусственного интеллекта (ИИ). В последнее десятилетие в мире было создано более 200 этических кодексов и рекомендаций, которые активно рассматриваются в рамках научных исследований. В основу большинства этих документов положен подход, основанный на принципах, задающий рамки этичности в сфере ИИ. В статье рассматриваются этические пробелы в этике искусственного интеллекта через разработку многоуровневой системы стратегий их преодоления. Предметом исследования выступают конкретные механизмы и стратегии трансформации абстрактных этических принципов в работающие практики на каждом из уровней. В данной статье рассматриваются основные подходы к решению проблем в сфере этики искусственного интеллекта. В качестве методологической основы предлагается системный подход, интегрирующий технические решения, организационные изменения и

правовые механизмы. Применимость предложенного подхода демонстрируется на анализе ключевых этических проблем ИИ: систематического воспроизведения предрассудков, «черного ящика» и распределения ответственности, декларативности этических и правовых норм, проблемы распределения ответственности. Для каждой проблемы выделены конкретные методы решения – от создания алгоритмов контрафактической справедливости до чек-листов проверки этичности систем ИИ. Сделаны выводы о том, что для системного устранения этических пробелов в этике ИИ необходимо создание трехуровневой системы «фильтров», дифференциациях подходов на уровне разработки, организационные изменения, создание новых принципов государственного регулирования и смена парадигмы в распределении ответственности. Основным выводом исследования является необходимость комплексного системного подхода, при котором строгость регулирования пропорциональна потенциальному вреду от использования системы ИИ.

Ключевые слова:

этика искусственного интеллекта, черный ящик, нейросети, этические кодексы, контрафактическая справедливость, принципы ИИ, этические пробелы, этика ИИ, ИИ-аудит, проблема интерпретируемости

Введение.

Исследование выполнено за счет гранта Российского научного фонда № 25-28-00254, <https://rscf.ru/project/25-28-00254>.

Развитие нейросетей и их активное внедрение в различные сферы экономики поставило задачу этической оценки и правового регулирования разработки и применения систем искусственного интеллекта (ИИ). В последнее десятилетие в мире было создано более 200 этических кодексов и рекомендаций (<https://oecd.ai/en/>). В основу большинства этих документов положен подход, основанный на принципах, задающий рамки этичности в сфере ИИ.

Этические кодексы активно исследуются для выявления сходства и различия подходов в различных странах и компаниях. Фундаментальные сравнительные исследования фиксируют, что в этических кодексах доминирует подход, основанный на принципах, наиболее часто встречаются принципы прозрачности, справедливости, ответственности, приватности [4]. Исследование Т. Хагендорфа, критикует декларативность и выявляет «слепые пятна» в этических кодексах, автор приходит к выводу о том, что этика ИИ страдает от институциональной слабости и декларативности, техническое сообщество демонстрирует отчуждение от этических вопросов, текущие руководства концентрируются на проблемах, допускающие технические исправления, но игнорируют широкий спектр социально-политических аспектов, поэтому необходим баланс между технологической конкретикой и социально-личностным измерением [2]. Последние исследования так же акцентируют внимание на декларативности юридических принципов в сфере ИИ [3]. Существуют исследования, показывающие, что на универсальные этические принципы в сфере ИИ оказывает влияние институциональный контекст [4]. Все эти группы исследований объединяет использование этических кодексов как материал для исследования

Поскольку большинство кодексов носят декларативный характер и фиксируют достаточно

общие установки и принципы, а опыт реального применения систем ИИ активно накапливается, важной задачей является определение этических и правовых пробелов. Так, Харитонова Ю.С., Савина В.С., Паньини Ф.^[5] в своем исследовании к основным пробелам в этике искусственного интеллекта относят декларативность этических и правовых норм, систематическое воспроизведение и усиление социальных предрассудков, непрозрачность принятия решений («черный ящик»), распределенный характер ответственности, недостаточность стандартизации и ряд других. В данной статье будут рассмотрены подходы к решению данных проблем.

Декларативность этических и правовых норм.

Одной из наиболее фундаментальных проблем регулирования сферы ИИ является пробел между формальным принятием этических кодексов и их реальным внедрением в практику разработки и бизнес-процессы. Устранение этического пробела, связанного с декларативностью этических и правовых норм кажется нам одним из наиболее трудоемких и требующих комплексного подхода.

Его преодоление возможно через создание многоуровневой системы «фильтров» взаимодействия с моделью. Таких уровней можно выделить три: уровень разработчиков, уровень заказчиков и государственный уровень. Если «фильтр» на первом или втором уровне пропустит неэтическое решение, то гарантией безопасности будет третий уровень, то есть институт государственной власти, который призван защищать и оберегать гражданские права, в том числе, и право на справедливость.

На уровне разработчиков целью и основной стратегией должно стать встраивание этических принципов непосредственно в технологию и процесс разработки. Для этого необходимо, с одной стороны, превращение этических принципов, таких как «справедливость», в конкретные математические метрики. С другой стороны, потребуется обучение разработчиков для формирования такой мировоззренческой установки, при которой этические вопросы при разработке встают в один ряд с техническими и юридическими. Данный путь требует достаточно высоких затрат для компании-разработчика.

Более простой вариант, который может дать результат – это *интеграция инструментов ИИ-аудита по аналогии с финансовым аудитом*. Примером может послужить интеграция SHAP-моделей. Еще одним вариантом решения может стать внедрение этических чек-листов для разработчиков (<https://disk.yandex.ru/i/YhC72pUrQICaBA>).

Риск-ориентированный подход как стратегия, согласно которой уровень строгости этического и правового контроля должен быть пропорционален потенциальному вреду от системы ИИ, позволяет сфокусировать ограниченные возможности регуляторов и разработчиков на самых критичных областях, таких как медицина, правосудие и финансовые услуги. Такой подход предотвращает избыточное регулирование в низко рисковых сферах, что позволяет создавать сбалансированную систему рисков, в которой требования к системе ИИ масштабируются в зависимости от потенциального воздействия на права и безопасность человека. Каждая из предложенных стратегий может работать в комплексе с другими, компания разработчик, таким образом, сможет пойти наиболее адекватным путем, исходя из собственных убеждений, стратегий и финансового положения.

На уровне заказчиков целью и основной стратегией становится изменение культуры и процессов внутри компании. Возможные меры: изменение структуры управления

включают добавление должностей «менеджера по ИИ», либо привлечение сторонних специалистов – аутсорсный ИИ-менеджмент, либо ИИ-аудит третьей стороной по аналогии с финансовым аудитом. Примером может послужить периодическая сертификация (раз в год) используемых систем искусственного интеллекта. Аудита и работы ИИ-менеджеров должны быть направлены на понимание преимуществ внедрения систем-ИИ и потенциальные правовые последствия от дискриминационных действий ИИ.

На третьем, государственном, уровне, акцент делается на создание правовых рамок, которые делают соблюдение этики обязательным, а не добровольным. Если первые два фильтра в виде разработчика и заказчика оказались недостаточно эффективными и/или отсутствующими, последний возможный рубеж – это государство – единственный общественный институт, имеющий право на законное принуждение. В своем регулировании государство может опираться на ряд принципов:

Принцип «доказанного соответствия» - компания должна доказать до выхода на рынок, что ее система ИИ соответствует стандартам безопасности. Например, беспилотный автомобиль не может быть выпущен на рынок, пока не будет доказана его полная безопасность.

Принцип «четкого определения ответственности» - законодательное закрепление цепочек ответственности за вред, причиненный ИИ. Например, дрон, управляемый ИИ разрушил дом, не являющийся объектом военной инфраструктуры, кто должен нести ответственность?

Определение «красных флагов» для принятия решений в рамках высокорисковых систем. Например, запрет на использование автоматизированных систем для принятия судебного решения в уголовном праве (ИИ-инструмент анализа, решение принимает судья).

Внедрение в законодательные акты «права на объяснение». Если отказано в страховом возмещении – гражданин имеет право на обоснованное и содержательное объяснение решения, принятого искусственным интеллектом.

Государственный уровень крайне необходим в данном случае, поскольку позволяет преодолеть декларативность, предложить реальные последствия – штрафы, ограничение деятельности, а также мотивировать к реальному внедрению этического поведения на двух других уровнях с помощью дотаций, ослабления налоговой нагрузки или льгот, предоставляемых компаниям, использующим ИИ.

Проблема систематического воспроизведения и усиления социальных предрассудков

Достаточно изученной является проблема систематического воспроизведения и усиления социальных предрассудков. Примером такой дискриминации может служить применение системы ИИ для отбора резюме кандидатов, когда система ИИ дискредитировала кандидатов по гендерному признаку [\[14, с. 87\]](#). В исследовании Джой Буоламини и Тимнит Гебру приводится анализ ошибок трех коммерческих систем распознавания лиц. По результатам исследования выяснилось, что ошибки распознавания женщин с темным цветом кожи являются наиболее неправильно идентифицируемой группой с частотой ошибок до 34,7%, в то время как частота ошибок для мужчин со светлой кожей обычно не превышает 0,8% [\[7\]](#). Кэти О'Нил приводит кейсы из образования, финансовой системы, как системы ИИ становятся «Орудием математического уничтожения» [\[8, с. 18\]](#). В отечественных исследованиях данная

проблема так же упоминается, например, можно указать на работы Ю.С.Харитонова, В.С.Савина [\[5\]](#), А.И.Смирновой [\[9\]](#), О.С.Бессараб , И.И.Газизова, А.Ф.Телякаева [\[10\]](#).

Можно обнаружить несколько вариантов решения предложенной этической проблемы. Первым приемом является *активный сбор репрезентативных данных*. Например, продуктом исследования Дж. Буолами и Т. Гебру является создание бенчмарка, представляющего собой сбалансированный набор изображений, размеченный по полу и типу кожи. Данный бенчмарк стал первым отраслевым стандартом для объективного сравнения и аудита коммерческих систем на предвзятость. Для построения модели необходим достаточный объем данных по всем социальным группам, что может быть затруднено недостатком или отсутствием информации. Решением проблемы может быть структурированный сбор данных до начала работы над обучением алгоритма. Второй вариант решения проблемы состоит во внедрении алгоритмов *контрфактической справедливости* [\[11\]](#) для отделения справедливых причин дискриминации от несправедливых. Например, при оценке того, что заемщик банка не выплатит кредит, для каждого заёмщика вычисляются вероятности невыполнения обязательств при изменении одного из параметров (пол, возраст, раса и т.д.). Ключевая проблема этого метода состоит в том, что он предъявляет высокие требования к данным и является чувствительным к пропущенным переменным. Это усложняет и удорожает процесс принятия решений, в связи с чем не все компании готовы жертвовать простотой ради этической строгости. Несмотря на сложности и ограничения использования данных методы, как правило, признаются весьма эффективными.

«Черный ящик» и проблема интерпретируемости.

Сложность моделей машинного обучения, особенно глубоких нейронных сетей, приводит к появлению феномена «черного ящика», когда процесс вычислений настолько непрозрачен, что даже эксперты не могут точно определить, что именно повлияло на то или иное решение системы ИИ. Для того, чтобы обеспечить доверие к системе ИИ, необходимо сделать ее легко интерпретируемой. При этом интерпретируемость должна быть доступна не только специалистам, «результаты и выводы системы должны быть представлены в доступной форме для широкого круга пользователей, включая неспециалистов» [\[12\]](#). Согласно ГОСТ Р 59276-2020 одним из факторов снижения качества системы ИИ называется «недостаточная понятность, объективность, предсказуемость и др.». В таком случае способом обеспечения доверия к системе ИИ должно служить «использование интеллектуальных алгоритмов обработки данных, обеспечивающих принятие системой объяснимых, предсказуемых и так далее решений» [\[13\]](#).

Решением проблемы можно назвать *использование более простых и интерпретируемых моделей в случаях, когда это возможно*. Например, сложная непрозрачная модель («черный ящик») отказала заемщику в предоставлении кредита. Заемщик не понимает отказ, снижается доверие к банку, банк рискует столкнуться с судебным иском за дискриминацию. В случае с интерпретируемой моделью система даёт объяснение того, на каких основаниях отклонена заявка (например, на основе дерева решений). Данный способ предполагает достаточно эффективное устранение пробела: легко понять почему модель ошиблась, внести корректировки, точечные исправления, аналитик может посмотреть все ветви дерева и проверить нет ли там правил, косвенно связанных с защищенными признаками. Проблема состоит в том, что обучение и тонкая настройка

простой модели может занять больше времени и ресурсов, чем использование готового продукта. Более того, модели «белого ящика» до сих пор считаются достаточно нишевой идеей [14, с. 376]. Другой вариант устранения пробела – *применение пост-хок анализа*, использование методов анализа для объяснения решений уже обученной модели черного ящика без необходимости понимания ее внутреннего устройства. Несмотря на сложности и ограничения применения метода: увеличенные требования к ресурсам и вычислительным мощностям, противоречивые ответы, получаемые при использовании разных методов, пост-хок анализ наиболее точно соответствует регуляторным требованиям, таким как право на объяснение, может применяться при отладке моделей и выявлении смещений и служит повышению доверия пользователей к системе.

Проблема распределения ответственности.

Системы ИИ, особенно те из них, что являются автоматизированными системами принятия решений, могут причинять вред различного рода, например, из-за предвзятости, связанной с дискриминацией, проблем с конфиденциальностью, безопасностью, манипулированию людьми при принятии решений. Часть из них, например, беспилотные летательные аппараты или беспилотные транспортные средства, может причинять конкретный физический вред, вплоть до лишения человека жизни. Группа авторов во главе с Б. Кастерсом выделяют четыре типа пробелов в ответственности:

1. Никто не может быть привлечен к ответственности (технология берет на себя процесс принятия решений).
2. Виновным признают не того, кого нужно (например, когда в случившемся обвиняют только непосредственного исполнителя).
3. Ответственность возлагается на нескольких участников (в конечном счете, они все избегают ответственности, перекладывая вину друг на друга).
4. Существует несколько правовых режимов (в таком случае, все участники избегают ответственности, ссылаясь на выгодный конкретно им правовой режим) [15].

Сам пробел существует на нескольких уровнях. Первый уровень определяется тем, что самообучающиеся алгоритмы функционируют не до конца предсказуемо даже для их создателей, а логика принятия ими решений может быть не интерпретируемой (проблема «черного ящика»). Второй уровень сложности состоит в распределении цепочки создания стоимости. В создании и использовании одной системы ИИ могут быть задействованы десятки сторон (разработчики алгоритмов, инженеры данных, владельцы и поставщики данных, производители процессоров, серверов, конечный оператор/пользователь, регулятор и т.д.), при этом каждая из сторон вносит свой вклад в конечный результат, а поэтому невозможно четко выделить, чье именно действие или бездействие стало решающим. Третий уровень – это общая правовая неопределенность. Например, правовая система оперирует понятиями «продавец», «изготовитель», «исполнитель», но система ИИ четко не вписывается ни в одну из категорий.

Рассмотрим стратегии преодоления пробела, основываясь на критическом подходе к этике искусственного интеллекта, предложенного Н.А. Ястреб [16]. Если объектом анализа является не сам ИИ, а гибридный комплекс акторов: начиная от самой капиталистической системой и заканчивая конечным пользователем, то вместо того,

чтобы пытаться «распределить» то, что по своей природе растворено, стоит предложить стратегию, основанную на принципах изменений. В таком случае, имеет смысл перейти от ответственности, как вины, к ответственности, как обязанности устранять ущерб. Зададим вопрос: «Кто извлекает основную выгоду из системы и обладает ресурсами для создания механизмов компенсации?» – вероятнее всего, правильным ответом станет оператор или владелец технологии, поэтому именно он и является первоочередным кандидатом для того, чтобы закрепить в отношении него *презумпцию обязанности по устранению ущерба*. Это не вина, а технико-юридическая обязанность, подобная обязанности энергокомпании устранять аварии на сетях, независимо от их причины. Подобная стратегия демифологизирует ИИ, переводя его из статуса «магического черного ящика», за который никто не отвечает» в статус «инженерной инфраструктуры», за функционирование которой кто-то отвечает по умолчанию.

Заключение

Статья посвящена анализу ключевых этических пробелов в этике искусственного интеллекта, возникающих на этапе активной технологической трансформации, вызванной внедрением систем искусственного интеллекта. Исследование показало, что несмотря на введение сотен этических кодексов и проработки принципов, они остаются во многом декларативными, что не позволяет говорить о реальном внедрении этических норм в практику. Основное противоречие современного этапа заключается в растущем разрыве между нормативными установками и их реализацией: технической и организационной.

Проведенный анализ позволил систематизировать четыре фундаментальных пробела в этике ИИ и предложить практико-ориентированные пути их преодоления. Во-первых, проблема декларативности норм требует создание многоуровневой системы контроля, включающей разработчиков, заказчиков и государство, где каждый уровень выступает в качестве «этического фильтра». Во-вторых, для решения проблемы систематического воспроизведения предрассудков, необходимы как технические меры (активный сбор репрезентативных данных, алгоритмы контрфактической справедливости), так и институциональные изменения. В-третьих, феномен «черного ящика» преодолевается через переход от поиска «виновного» к модели «обязанности устраниТЬ ущерб», где основное бремя ложится на оператора технологии.

Ключевым выводом исследования является необходимость комплексного и риск-ориентированного подхода, при котором строгость регулирования пропорциональна потенциальному вреду от системы ИИ. Особое значение имеет государственное регулирование, создающее обязательные правовые рамки, включающие принципы доказанного соответствия, четкого определения ответственности и права на объяснение. Однако эффективность регулирования зависит от параллельного развития этической культуры на уровне разработчиков и заказчиков.

Библиография

1. Jobin A., Ienca M., Vayena E. The Global landscape of AI ethics guidelines // Nature Machine Intelligence. 2019. Vol. 1, No. 9. P. 389-399.
2. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines. 2020. P. 99-120.
3. Morin-Martel A. AI Ethics: Institutional Turn // Digital Society. 2025. No. 4(18).
4. Bar-Gil O. Examining trends in AI ethics across countries and institutions via quantitative discourse analysis // AI & Society. 2025.
5. Харитонова Ю. С., Савина В. С., Паньини Ф. Предвзятость алгоритмов искусственного

- интеллекта: вопросы этики и права // Вестник Пермского университета. Юридические науки. 2021. Вып. 53. С. 488-515. DOI: 10.17072/1995-4190-2021-53-488-515. EDN: EUKCPY.
6. Шталь Б. К., Шредер Д., Родригес Р. Этика искусственного интеллекта: кейсы и варианты решения этических проблем // Экономическая социология. 2024. Т. 25. № 1. С. 85-92. DOI: 10.17323/1726-3247-2024-1-85-95. EDN: HVDDTQ.
7. Buolomwini J., Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification // Proceedings of Machine Learning Research. 2018. Р. 1-15.
8. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016. 209 р.
9. Смирнова А. И. Предвзятость как проблема алгоритмов ИИ: этические аспекты // Философия и общество. 2023. № 3. С. 118-126. DOI: 10.20884/jfio/2023.03.07. EDN: VNZKTC.
10. Бессараб О. С., Газизова И. И., Телякаева А. Ф. Философия искусственного интеллекта и этика технологий: как философия формирует этические принципы для развития ИИ, моральные дилеммы, связанные с роботизацией // Вестник науки. 2025. № 6(87). Т. 1. С. 1285-1296. EDN: JDJW XV.
11. Чесалов А. Ю. Глоссариум по искусенному интеллекту: 2500 терминов. М.: Издательские решения, 2022. 670 с.
12. Барышников П. Н. "Черный ящик" или прозрачный алгоритм: аналитический обзор источников по этике искусственного интеллекта // Вестник Томского государственного университета. Философия. Социология. Политология. 2025. № 85. С. 5-20. DOI: 10.17223/1998863X/85/1. EDN: IWICPO.
13. ГОСТ Р 59276-2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения [Электронный ресурс] // Федеральное агентство по техническому регулированию и метрологии (Росстандарт). Режим доступа: <https://www.rst.gov.ru/portal/gost> (дата обращения 01.12.2025).
14. Бирюков Д. Н., Дудкин А. С. Объяснимость и интерпретируемость – важные аспекты безопасности, 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.
15. Custers B., Lahmann H., Scott B. From liability gaps to liability overlaps: shared responsibilities and fiduciary duties in AI and other complex technologies // AI & Society. 2025. Р. 4035-4050.
16. Ястреб Н. А. Концепции этики искусственного интеллекта: от принципов к критическому подходу // Семиотические исследования. 2024. Т. 4. № 1. С. 24-30. DOI: 10.18287/2782-2966-2024-4-1-24-30. EDN: UFLHKD.
17. Ястреб Н. А. Идеализированная когнитивная модель абстрактного понятия "искусственный интеллект" // Вестник Томского государственного университета. Философия. Социология. Политология. 2025. № 84. С. 44-54. DOI: 10.17223/1998863X/84/5. EDN: SJPHNJ. ""

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](#).

Предметом рецензируемой статьи являются этические пробелы в этике искусственного

интеллекта. В самом широком смысле под искусственным интеллектом понимается интеллект машин или компьютерных систем, созданный усилиями человеческого гения. На наш взгляд, этическая составляющая проблемы искусственного интеллекта тесно связана с осознанием особенностей мыслительных процессов самого человека, ибо целью создания ИИ является достижение машиной человеческого уровня мышления или даже более высокой степени, нежели у человека. Все ошибки и возможные «пробелы» в этике ИИ генетически связаны с обществом, создавшим этот интеллект. Все социальные болезни, присущие капитализму как общественно-экономической формации, неизбежно отражаются в ИИ. Автор отмечает, что статья посвящена анализу ключевых этических пробелов в этике искусственного интеллекта, возникающих в процессе активной технологической трансформации, происходящей вследствие внедрения систем искусственного интеллекта. Автору рекомендуется более четко сформулировать объект и предмет исследования и сделать это не в заключении, а в начале статьи. Хотелось бы также, чтобы автор подумал над возможностью избежать тавтологии в заглавии статьи («этические пробелы в этике»). Можно предложить заменить эту формулировку на «нравственные пробелы». Среди таких «пробелов» автор, основываясь на обзоре научной литературы, справедливо выделяет декларативность этических и правовых норм, систематическое воспроизведение и усиление социальных предрассудков, непрозрачность принятия решений («черный ящик»), распределенный характер ответственности, недостаточность стандартизации.

Статья носит, по большей части, обзорный характер, и методология исследования основана на этической и социально-философской компаративистике. Автор рецензируемого исследования использует также категориальный анализ и дискурсивный подход. Статья демонстрирует основательное знакомство автора с имеющейся научной литературой по исследуемой им проблематике.

Рецензируемая статья представляет собой актуальное исследование, ибо дискуссии об искусственном интеллекте, его сущности, возможностях и потенциальных опасностях, таящихся в нем для человечества, стали вестись и в философских сообществах – пусть и с явным опозданием. Хотя плодотворное начало этим дискуссиям положили великие писатели-фантасты А. Азимов, Р. Шекли, Р. Брэдбери и другие мэтры. Автор совершенно справедливо отмечает, что «развитие нейросетей и их активное внедрение в различные сферы экономики поставило задачу этической оценки и правового регулирования разработки и применения систем искусственного интеллекта». Здесь проблемы этики как части философии выходят на сугубо практические задачи, связанные с выживанием человека в новой технократической реальности.

Статья, безусловно, обладает определенной исследовательской новизной. Рецензент согласен с выводом автора, что декларативность большинства этических принципов не позволяет внедрить их в полной мере в практику. Автор видит основное противоречие, возникающее при использовании ИИ, в растущем разрыве между нормативными установками и их технической и организационной реализацией. Проведенный автором анализ позволил систематизировать четыре фундаментальных пробела в этике ИИ и предложить практико-ориентированные пути их преодоления. Но в заключении статьи, перечисляя эти пробелы, автор говорит лишь о трех. Хотелось бы большей четкости в этом аспекте.

Стиль статьи в полной мере соответствует стилю научной работы, статья познавательна по своему содержанию, а изложенные в ней факты и выводы заставляют задуматься не только теоретиков, но и практиков, занимающихся внедрением ИИ. Структура исследования вполне логична и способствует выражению основных идей автора.

Библиография статьи включает важные работы, относящиеся к теме исследования, хотя рецензент и надеялся обнаружить в этом списке работы Стюарта Рассела и Петера

Норвига, или же Алексея Потапова. Но это – дело вкуса, да и работ об ИИ скоро будет написано больше, чем было написано о Шекспире.

Рецензируемая статья, на наш взгляд, представляет собой исследование на тему, ставшую за последние годы особенно актуальной и многоаспектной. Автор прав, когда пишет, что эффективность государственного регулирования зависит от параллельного развития этической культуры на уровне разработчиков и заказчиков. На наш взгляд, многие мысли автора стоят того, чтобы быть продолженными в его дальнейших исследованиях. Данная статья, безусловно, заслуживает публикации после устранения указанных рецензентом замечаний и неточностей.