

УДК 004.8

## ***ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ***

### ***ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ***

***Салова Т.Л.***

*кандидат технических наук, доцент,*

*кафедра Информационных технологий и математики,*

*Сочинский государственный университет,*

*Сочи, Россия*

***Ковгореня М.В.***

*Магистрант,*

*Сочинский государственный университет,*

*Сочи, Россия*

#### **Аннотация**

В статье рассматриваются этические и правовые аспекты внедрения технологий искусственного интеллекта в условиях трансформации современного информационного общества. Показано, что темпы развития искусственного интеллекта опережают формирование нормативно-правовой базы и этических рамок, что повышает риски нарушения прав человека, дискриминации и снижения доверия к цифровым сервисам. Основное внимание уделено проблемам непрозрачности алгоритмов («чёрного ящика»), обеспечению объяснимости результатов, а также механизмам подотчётности и аудита систем искусственного интеллекта. Проведён анализ существующих регуляторных подходов и практик «мягкого» регулирования (этические кодексы, отраслевые стандарты) в России и на международном уровне, включая экспериментальные правовые режимы. Обоснована необходимость разграничения сфер допустимой автоматизации и областей, требующих обязательного человеческого контроля. Предложены

направления развития системы ответственности разработчиков и операторов искусственного интеллекта, ориентированные на повышение прозрачности, управление рисками и защиту прав пользователей.

**Ключевые слова:** искусственный интеллект, этика искусственного интеллекта, прозрачность алгоритмов, права пользователей, общий регламент по защите данных, прикладная информатика, информационное общество, алгоритмическая ответственность.

## ***ETHICAL AND LEGAL ASPECTS OF THE USE OF ARTIFICIAL INTELLIGENCE IN THE INFORMATION SOCIETY***

***Salova T.L.***

*Candidate of Technical Sciences, an associate Professor,  
Department of Information Technologies and mathematics  
Sochi State University,  
Sochi, Russia*

***Kovgorenya M.V.***

*Undergraduate,  
Sochi State University,  
Sochi, Russia*

### **Abstract**

The article discusses the ethical and legal aspects of the introduction of artificial intelligence technologies in the context of the transformation of the modern information society. It is shown that the pace of development of artificial intelligence is ahead of the formation of a regulatory framework and ethical framework, which increases the risks of human rights violations, discrimination and a decrease in trust in digital services. The main attention is paid to the problems of non-transparency of algorithms ("black box"),

ensuring the explainability of results, as well as mechanisms for accountability and auditing of artificial intelligence systems. The analysis of existing regulatory approaches and practices of "soft" regulation (codes of ethics, industry standards) in Russia and at the international level, including experimental legal regimes, is carried out. The need to distinguish between the spheres of permissible automation and the areas requiring mandatory human control is substantiated. Directions for the development of the responsibility system of developers and operators of artificial intelligence, focused on increasing transparency, risk management and protecting the rights of users, are proposed.

**Keywords:** artificial intelligence, artificial intelligence ethics, algorithm transparency, user rights, general data protection regulations, applied informatics, information society, algorithmic responsibility

**Введение.** Современное информационное общество переживает глубокую трансформацию, связанную с внедрением технологий искусственного интеллекта во все сферы человеческой деятельности. Стремительное развитие этих технологий опережает формирование адекватной нормативно-правовой базы и этических рамок. Данная статья исследует ключевые этические и правовые аспекты применения ИИ в информационном обществе, с особым акцентом на вызовы, которые стоят перед прикладной информатикой в обеспечении прозрачности алгоритмов и защиты прав пользователей в России и на международном уровне. Анализируются существующие регуляторные подходы, проблемы дискриминации и предвзятости, а также предлагаются пути развития системы ответственности разработчиков и операторов ИИ-систем [1,2].

**Обсуждение и результаты.** Информационное общество представляет собой качественно новый этап развития человечества, характеризующийся центральной ролью информации и знаний в экономической и социальной деятельности. Цифровизация затрагивает все сферы жизни: государственное управление, Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ Эл № ФС 77-68405 ISSN 2541-8327

здравоохранение, образование, финансовый сектор, судебную систему и правоохранительные органы. Технологии искусственного интеллекта (ИИ) открывают беспрецедентные возможности для анализа больших данных, автоматизации процессов и решения сложных социальных проблем. Внедрение ИИ без надлежащих этических и правовых рамок может привести к усугублению существующего неравенства, нарушению прав человека и подрыву доверия к цифровым технологиям [3].

Цифровизация несёт в себе двойственный характер угроз и возможностей. Она может стать катализатором устойчивого развития и решения глобальных проблем, но при этом беспрепятственное развитие технологий может привести к распространению фейковых новостей, появлению цифрового авторитаризма и усилению власти крупных технологических компаний. Государства перед лицом этих вызовов стоят перед двойной задачей: раскрыть огромный потенциал новых цифровых технологий и одновременно предотвратить разрушительные последствия их неконтролируемого развития [4]. Это требует формирования комплексной системы регулирования, которая балансировала бы интересы инноваций, безопасности и защиты основных прав человека [4, 5].

На международном уровне действует стандарт ЮНЕСКО: «Рекомендация об этических аспектах искусственного интеллекта» [11], который появился в 2021 году и носит рекомендательный характер для стран – членов ООН. Основным разработчиком стандартов этики ИИ в настоящий момент можно считать Международную организацию по стандартизации (ISO). Правда, следует отметить, что в США нет федерального закона об ИИ, а регулирование строится на добровольной основе и инициативе штатов, в отличие от Китая, который предлагает централизованную модель регулирования. Этика искусственного интеллекта представляет собой комплекс рекомендательных принципов и правил, предназначенных для создания среды доверенного развития технологий ИИ и

Дневник науки | [www.dnevnikaui.ru](http://www.dnevnikaui.ru) | СМИ Эл № ФС 77-68405 ISSN 2541-8327

существенно зависит от согласованности правил и создания общей инфраструктуры доверия международного сообщества [12, 13].

В России этика ИИ получила своё концептуальное воплощение в Кодексе этики в сфере искусственного интеллекта (26 октября 2021 года) [6]. Этот документ носит рекомендательный характер и закрепляет общие этические принципы и стандарты поведения для лиц, разрабатывающих, применяющих или эксплуатирующих системы ИИ. Основные принципы включают приоритет защиты интересов людей и тезис о том, что ответственность за последствия применения ИИ лежит на человеке.

Кодекс этики в сфере разработки и применения искусственного интеллекта на финансовом рынке (Банк России, 2025) конкретизирует эти принципы для финансового сектора [1]. Он основан на пяти ключевых принципах: человекоцентричность, справедливость, прозрачность, безопасность и ответственное управление рисками. Такое «мягкое» регулирование позволяет задавать стандарты поведения и практики без введения отдельного жесткого «рамочного закона» для всей сферы.

Внедрение ИИ создаёт фундаментальные этические дилеммы, которые требуют глубокого анализа. При наличии мощных систем автоматизации принятия решений человек рискует оказаться перед принуждением подчиняться логике алгоритма, лишаясь возможности опираться на контекстное человеческое понимание. Принципы права — гуманизм, равенство и справедливость — могут оказаться под давлением при замене человеческих решений автоматизированными.

Решение этой дилеммы требует разграничения сфер, где применение ИИ допустимо, и сфер, где необходимо сохранить человеческое суждение и контроль. Кодекс этики РФ устанавливает приоритет защиты интересов людей как базовый ориентир разработки и применения ИИ [1]. Это означает, что внедрение ИИ должно

оцениваться с точки зрения его влияния на достоинство, автономию и благополучие человека [6].

На настоящий момент в России ключевым стратегическим документом является Указ Президента РФ № 490 (10.10.2019), который задаёт направление развития ИИ и рамку государственных мер [9]. В отсутствие единого специального федерального закона об ИИ значимую роль играют также этические кодексы и режимы апробации технологий [6].

С 1 июля 2020 года в Москве вводится экспериментальный правовой режим в сфере применения технологий искусственного интеллекта [7]. Основными принципами режима являются прозрачность, защита прав и безопасность, а также недискриминационный доступ к результатам применения ИИ. Этот механизм предназначен для накопления практики и последующей донастройки регулирования.

Одна из центральных этических проблем, связанных с применением ИИ, — непрозрачность алгоритмов («чёрный ящик») [6]. Когда алгоритмы остаются непонятными для пользователей и затрудняют проверку результатов, это снижает доверие и усложняет аудит. В сфере высокого риска (например, финансы и иные социально значимые области) непрозрачность может привести к неправомерным решениям, дискриминации или нарушению конфиденциальности [5].

Прозрачность ИИ становится не только технической задачей, но и условием подотчётности и правовой защищённости пользователя. Для этого применяются подходы Explainable AI (XAI), позволяющие повышать интерпретируемость и объяснимость результатов [8].

Объяснимость и интерпретируемость — связанные, но различные концепции. Интерпретируемость — это свойство модели, при котором её логика работы понятна человеку напрямую, без дополнительных инструментов. Такая модель устроена так, что можно проследить, как именно входные данные преобразуются в

Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ Эл № ФС 77-68405 ISSN 2541-8327

результат: какие признаки влияют на решение, с каким весом, по каким правилам. Типичные примеры — линейная регрессия, дерево решений или система чётких правил «если пользователь видит механизм принятия решения и может его осмыслить без посредников.

Объяснимость, напротив, не требует, чтобы модель была проста по структуре. Даже сложный «чёрный ящик» (например, глубокая нейронная сеть) может быть объяснимым — если он способен генерировать понятные человеку пояснения к своим решениям. Эти пояснения создаются уже после работы модели с помощью специальных методов (таких как LIME или SHAP) и могут принимать разные формы: текстовые объяснения, визуализации, выделение значимых фрагментов входных данных. Главное — результат адаптирован под восприятие человека и отвечает на вопрос «почему было принято именно это решение?».

Хотя обе концепции служат одной цели — сделать ИИ прозрачными и доверительными, — они решают эту задачу разными путями. Интерпретируемость делает ставку на внутреннюю прозрачность: модель сама по себе должна быть понятна. Объяснимость же фокусируется на внешней коммуникации: даже непрозрачная модель может «объяснить» свой вывод так, чтобы человек его понял.

Это различие проявляется и в практических ограничениях. Интерпретируемые модели зачастую проигрывают в точности более сложным аналогам, поскольку их простота накладывает ограничения на выразительность. Зато их логика всегда на виду. Объяснимые же системы могут достигать высокой точности за счёт сложной архитектуры, но их объяснения — это всегда некий «перевод» с языка модели на язык человека, а не прямое отображение внутренних процессов. Эффективная прозрачность включает: информирование пользователя о

применении ИИ, возможность получить объяснение результата и наличие процедур контроля/аудита [8, 9].

При реализации принципов прозрачности возникает противоречие: с одной стороны — потребность общества в открытости, с другой — необходимость сохранения коммерческой тайны разработчиков [10]. Поэтому в качестве практических решений рассматриваются независимый аудит, стандарты раскрытия для высокорисковых систем и организационные меры управления рисками [2]. В финансовом секторе соответствующие принципы закреплены в кодексе Банка России.

Персональные данные являются критически важным ресурсом для обучения ИИ-систем, но их обработка усиливает требования к правовой защите и управлению рисками [6]. При применении ИИ в социально значимых сферах особенно важны принципы человекоцентричности, прозрачности и безопасной эксплуатации.

Исследования и прикладные кейсы показывают, что ИИ-системы могут воспроизводить и усиливать дискриминацию из-за смещения в данных и в процедурах обучения. Поэтому для высокорисковых сценариев нужны процедуры тестирования на справедливость мониторинга и внешней проверки, включая аудит алгоритмов [4].

Одной из центральных проблем правового регулирования ИИ является распределение ответственности между разработчиком, владельцем, оператором и пользователем. Для обеспечения ответственности необходимо развитие систем аудита алгоритмов (аналогично финансовому аудиту) и процедур документирования решений ИИ. На практике отраслевые кодексы также фиксируют необходимость ответственного управления рисками.

**Заключение.** В условиях современного информационного общества стремительное внедрение технологий искусственного интеллекта заметно меняет

Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ Эл № ФС 77-68405 ISSN 2541-8327

экономику, государственное управление, образование, медицину и повседневные практики. Однако темпы технологического развития часто опережают становление полноценной нормативно-правовой базы и согласованных этических ориентиров. В результате усиливаются риски дискриминации и предвзятых решений, нарушения прав и свобод человека, а также снижения общественного доверия к цифровым сервисам.

Среди ключевых вызовов выделяются обеспечение прозрачности и объяснимости алгоритмов, надёжная защита прав пользователей и персональных данных, а также справедливое распределение ответственности между разработчиками, владельцами, операторами и конечными пользователями ИИ-систем. Дополнительно актуализируются вопросы независимой оценки качества моделей, аудита данных и механизмов контроля, предотвращающих злоупотребления и небезопасное применение ИИ в критически важных сферах.

С целью оценки этичности и прозрачности работ систем ИИ целесообразно использовать хорошо зарекомендовавший себя на международном уровне механизм алгоритмического аудита [14].

### **Библиографический список**

1. Кодекс этики в сфере разработки и применения искусственного интеллекта на финансовом рынке [Электронный ресурс] // Официальный сайт Банка России. — 2025. — URL: [https://cbr.ru/Content/Document/File/178667/code\\_09072025.pdf](https://cbr.ru/Content/Document/File/178667/code_09072025.pdf) (дата обращения: 23.11.2025)

2. Буряга В.О., Джужома В.В., Артеменко Е.А. Формирование модели регулирования искусственного интеллекта: международный и отечественный опыт // Правовые вопросы в цифровую эпоху. 2025. Т. 6. № 2. С. 50–68.

3. Головин К.С. Необходимость правового регулирования «черного ящика» искусственного интеллекта // Вестник Костромского государственного университета. 2024. Т. 30. № 3. С. 290–297.

4. Дудикова Е.И., Куницына Н.Н. Поляризация информационного общества: цифровой перезагрузка // Вестник РУДН. Серия: Социология. 2024. Т. 24. № 2. С. 539–554.
5. Жарова А.К. Достижение алгоритмической прозрачности и управление рисками информационной безопасности при принятии решений без вмешательства человека: правовые подходы // Journal of Digital Technologies and Law. 2023. № 1(4). С. 973–993.
6. Кодекс этики в сфере искусственного интеллекта (от 26 октября 2021 г.) — [Электронный ресурс] — URL: <https://ethics.a-ai.ru> (дата обращения: 23.11.2025).
7. Объяснимые нейросети (ХАИ): почему ИИ должен быть прозрачным? — [Электронный ресурс] — URL: <https://habr.com/ru/companies/timeweb/articles/930764/> (дата обращения: 25.11.2023).
8. Талапина Э.В. Прозрачность алгоритмов искусственного интеллекта // Право. Журнал Высшей школы экономики. 2025. Том 18. № 3, с 4-27
9. Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации»
10. Федеральный закон от 24.04.2020 №123-ФЗ // Официальный интернет-портал правовой информации. — 2020. URL: <https://base.garant.ru/73945195/> (дата обращения: 25.11.2023).
11. ЮНЕСКО. Рекомендация об этических аспектах искусственного интеллекта. 2021, с.1-39.
12. Keng Siau, Weiyu Wang. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. March 2020. Journal of Database Management 31(2), p.74-87.
13. Machine Ethics /ed. by M. Anderson, S. Anderson. – New York; Cambridge: University Press, 2021. – 548 p.

14. Namiot D., Ilyushin E. Trusted Artificial Intelligence Platforms: Certification and Audit. // International Journal of Open Information Technologies. ISSN: 2307-8162. Vol.12, no.1, 2024.