

Административное и муниципальное право

Правильная ссылка на статью:

Атабеков А.Р. — Обеспечение автономности принятия решения искусственным интеллектом для целей публичных правоотношений // Административное и муниципальное право. – 2023. – № 1. DOI: 10.7256/2454-0595.2023.1.39893 EDN: GYFCBG URL: [https://nbppublish.com/library\\_read\\_article.php?id=39893](https://nbppublish.com/library_read_article.php?id=39893)

## Обеспечение автономности принятия решения искусственным интеллектом для целей публичных правоотношений

Атабеков Атабек Рустамович

кандидат экономических наук

доцент, кафедра административного и финансового права, Российский университет дружбы народов

117198, Россия, г. Москва, ул. Миклухо-Маклая, 6

✉ atabekoff1@mail.ru



[Статья из рубрики "АДМИНИСТРАТИВНОЕ И МУНИЦИПАЛЬНОЕ ПРАВО И ПРОБЛЕМЫ ИНФОРМАТИЗАЦИИ"](#)

### DOI:

10.7256/2454-0595.2023.1.39893

### EDN:

GYFCBG

### Дата направления статьи в редакцию:

27-02-2023

### Дата публикации:

06-03-2023

**Аннотация:** Предметом исследования выступает формализация действий искусственного интеллекта в качестве представителя органа власти. Объектом исследования выступают нормативные документы, рекомендации и иные документы регламентирующие вопросы реализации автономности искусственного интеллекта для целей публичных правоотношений в России и зарубежных странах, судебная практика, академические публикации и аналитические отчеты по исследуемой проблематике. Методология исследования интегрирует комплекс современных философских, общенациональных, специально-научных методов познания, включая, диалектический, системный, структурно-функциональный, герменевтический, сравнительно-правовой, формально-юридический (догматический) и др. В рамках настоящего исследования делается особый акцент на осуществлении сравнительного правового исследования феномена автономности искусственного интеллекта, реализующего публичные функции на основании опыта различных государств. В рамках указанной статьи проводится сравнительный анализ действующих подходов по определению базовых условий

обеспечения автономности искусственного интеллекта в контексте публичных правоотношений зарубежных стран и России. В рамках проведенного сравнительного анализа выявлены базовые проблемы в области прозрачности принятия решения искусственного интеллекта в мировой практике, определены практические ситуации интеграции непрозрачных искусственного интеллекта в сферу публичных правоотношений зарубежных стран, а также предложены возможные компенсирующие правовые мероприятия, обеспечивающие безопасную интеграцию искусственного интеллекта в сферу государственного управления в России. Предложенные в результате проведенного исследования мероприятия могут найти свое применение в законодательной и правоприменительной практике профильных органов власти реализующих интеграцию искусственного интеллекта в сферу общественных и публичных отношений в России.

**Ключевые слова:**

искусственный интеллект, электронное лицо, сравнительно-правовое исследование ИИ, машинное обучение, контрафактуальный анализ, безопасный ИИ, публичное право, административное право, информационное право, правоприменительная практика

Регламентация правоотношений, возникающих из феномена искусственного интеллекта и прозрачности принятия решений для целей публичных правоотношений, является стратегически значимым как для России, так и для зарубежных стран. При этом, отдельную роль занимает вопрос позиционирования ИИ в сфере публичных правоотношений, остается системным и комплексным вопрос как на уровне теоретических исследований, так и практических решений в области понятности и прозрачности принятых ИИ решений.

На уровне Министерства юстиции Германии<sup>[1]</sup>, Комиссии ЕС<sup>[2]</sup>, отдельных штатов США<sup>[3]</sup> уже системно рассматривается вопрос необходимости разрешения «черного ящика».

Среди представителей зарубежной и отечественной научной школы также системно возникают вопросы как в части технической осуществимости раскрытия этого «ящика»<sup>[4]</sup>, а также того, что иллюзия подотчетности ИИ для человека поднимает вопрос терминологической определенности понятий «объяснимости», «постижимости», «космысленности» и т.д.<sup>[5-8]</sup>.

При рассмотрении технико-правового аспекта функционирования ИИ для целей публичной власти первым аспектом является обеспечение прозрачности ИИ, поскольку используется динамичная структура обрабатываемых данных, которая выражается отсутствием прямой взаимосвязи входных данных и «выходных» решений (итогов)<sup>[9]</sup>. Причина кроется в том, что ИИ (в особенности, базирующиеся на базе машинного обучения) используют большое количество способов анализа и достижения цели имеющие различные алгоритмические принципы и подходы к их интерпретации<sup>[10]</sup>. Например, вопрос интерпретации машинописного текста, логарифмическая формула оценки ущерба (предусмотренная тем или иным подведомственным учреждением), доказательная база стороны участника административного процесса, которая может включать устные пояснения с использованием нестандартных речевых оборотов и т.д. и обработка данных ИИ.

Кроме того, вопрос скорости обработки данных и постоянная обработка данных формирует условно «живой организм», имеющий постоянную динамическую структуру данных, потенциально обновляющуюся при каждом взаимодействии с пользователем [11-[12](#)].

При этом на уровне доктринального восприятия ИИ для целей публичных правоотношений необходимо отметить так же, что ИИ и его базисный составной элемент - машинное обучение- работают по той выборке дел, которые ранее были наработаны соответствующим должностным лицом.

Следует понимать, что общая кодификация административного права и процесса позволяет в большинстве своем поставить деятельность служащих профильных министерств и ведомств на «алгоритмические рельсы»; в то же время нельзя не отметить, то, что в рамках каждого дела присутствуют как общные черты (присущие тому или иному классу дел, предусмотренных кодексами), так и отличительные черты, формирующие подмножество случаев с обобщаемыми свойствами.

Необходимо понимать, что вопрос обобщения административных дел, равно как и их персонализация для целей ИИ, порождает ряд проблем, таких как проблемы эффекта отбора и сформированной предвзятости.

Разбирая проблему эффекта отбора, мы видим, что алгоритм работающий на одной базе данных, может перестать эффективно реализовывать свой публичный функционал для целей другого ведомства или же генерации данных в том же сегменте, но со случайной выборкой исходных данных [13]. Вопрос предвзятости данных определяется тем, насколько применяемая норма используется в административной практике и есть ли сформированная репрезентативная выборка фактов, что так же систематически выделяется как отечественными, так и зарубежными учеными [14-16].

Базовым решением указанных проблем могут выступать три следующих инструмента (по отдельности или в связке):

1. Генерация случайных ситуаций/выборок, которая позволяет наиболее точно формулировать практику выработки решений ИИ, в особенности, если базой данной выступает наиболее широкий круг дел, принимаемых различными должностными лицами в рамках одного процесса [17].
2. Применение контрфактального анализа дела [18], который дает генерацию ответвлений на принятые решения, и сегменты, влияющие на принятие решений ИИ; информационный шум, создаваемый данной выборкой и ответвлений минимизирует число нешаблонных решений, принятых человеком.
3. Введение связки ИИ + человек, где ИИ выступает средством мониторинга случайных выборок, выделяющих проблемные зоны должностному лицу для верификации и оценки доказательств.

Следует понимать, что указанная связка тоже может сохранять риски неэтичного принятия решения, недостоверности выборки при принятии решения ИИ (человеческий фактор), а также потенциальное влияние ИИ на базовые права, предусмотренные Конституцией (ч.2 ст. 19). Однако мы понимаем, что данные действия позволят в последствии сделать человеческие предубеждения или ошибки более заметными.

Кроме того, нужно учитывать, что технология ИИ может быть использована не только во

благо государства, но и в противовес действующим мерам, предусмотренным законодательства.

Весь комплекс заложенных мер при принятии автономного решения ИИ может быть эмпирически идентифицирован, и, как следствие, могут быть сформулированы стратегические контрмеры, предусмотренные реализацией принципов правовой определенности [\[19\]](#).

Среди таковых можно выделить вопросы создания информационного шума используемых баз данных ИИ (в особенности находящихся за контуром формирования профильным органов власти), а также формирование поведенческой адаптации, когда подконтрольный субъект, находящийся вне риска попадания в выборку, менее «законопослушен».

Данные доктринальные проработки вопросов интеграции ИИ в сферу публичных правоотношений находят свое отражение в практической публичной деятельности органов власти различных стран:

Власти Нидерландов в рамках контроля мошенничества в сфере социального обеспечения использовали систему SyRi [\[20\]](#), которая, по сути, использовала в своей работе большой массив данных, но при этом не была достаточно прозрачной в своей обработке данных и предоставляемых эмпирических результатах. В результате указанный алгоритм был запрещен к применению органом власти ввиду несоответствия его принципов статье 8 ЕКПЧ (Европейской конвенции о правах человека) [\[21\]](#).

Министерство труда и социальной политики Польши использовало узкоспециализированный ИИ, который должен был проводить категоризацию потенциальных получателей пособий по безработице [\[22\]](#). При этом бинарная обработка решения ИИ проходила соответствующую верификацию сотрудников ведомства, однако по причине халатного отношения в почти 100% случаев позиция ИИ не оспаривалась служащим. В результате анализа действий ИИ на предмет прозрачности, Конституционный суд Польши признал данный продукт не конституционным [\[23\]](#).

Агентство по страхованию от безработицы штата Мичиган (США), использовало аналогичный SyRi инструмент для целей выявления случаев мошенничества, связанных с получением пособия по безработице. При этом указанный ИИ осуществлял полностью автономное принятие решения в отношении взыскания пособия, без возможности апелляции указанного решения в рамках ведомства [\[24\]](#), что в последствии повлекло судебное разбирательство и соответствующие штрафные санкции к разработчику ИИ [\[25\]](#).

На основании вышеизложенного, в целях формирования взвешенного подхода по применению ИИ для целей публичных правоотношений предлагается следующее:

1. Необходимо определение секторов применения ИИ, которые не влекут за собой нарушение базовых конституционных прав граждан.
2. Необходимо формирование тестовых полигонов данных в рамках профильных ФОИВов, с возможностью оперативного реагирования в рамках сбоев, ошибок и т.д. допускаемых ИИ и профильных сотрудников ФОИВов.
3. Необходимо применение ИИ на наиболее репрезентативной выборке дел, исключающей риск возникновения предсказанного преследования ввиду низкого

- качества выборки дел или слабой административной практики в рамках указанного сектора правоотношений.
4. Необходимо внедрение контрафактуальных объяснений ИИ в действиях всех участников процесса, в целях снятия проблем, обусловленных технологической природой ИИ.
  5. Необходим контроль рисков, связанных с ИИ, в контексте формирования контрмер со стороны злоумышленников, посредством создания информационных шумов во внешних базах данных или формирования поведенческой адаптации у целевой категории граждан.
  6. При внедрении полностью автономных ИИ важно учитывать сложившийся негативный международный опыт в части не прозрачных и неконституционных механизмов принятия решения ИИ, с внедрением в российском правоприменительном поле соответствующих компенсационных механизмов обеспечения прозрачности принятия решения и его верификации профильным сотрудником ФОИВа.

## **Библиография**

1. Zwischenbericht der Arbeitsgruppe "Digitaler Neustart" zur Frühjahrskonferenz der Justizministerinnen und Justizminister am 6. und 7. Juni 2018 in Eisenach: [сайт]. — URL: [www.justiz.nrw.de/JM/schwerpunkte/digitaler\\_neustart/zt\\_fortsetzung\\_arbeitsgruppe\\_teil\\_2/2018-04-23-Zwischenbericht-F-Jumiko-2018%2D%2D-final.pdf](http://www.justiz.nrw.de/JM/schwerpunkte/digitaler_neustart/zt_fortsetzung_arbeitsgruppe_teil_2/2018-04-23-Zwischenbericht-F-Jumiko-2018%2D%2D-final.pdf) (дата обращения: 21.02.2023).
2. Proposal for a Regulation on promoting fairness and transparency for business users of online intermediation services (COM(2018) 238 final / 2018/0112 (COD)): [сайт]. — URL: [https://eur-lex.europa.eu/procedure/EN/2018\\_112](https://eur-lex.europa.eu/procedure/EN/2018_112) (дата обращения: 21.02.2023).
3. The initial proposal (Int. 1696–2017) would have added the text cited above to Section 23-502 of the Administrative Code of the City of New York. However, the law that was finally passed only established a task force which is designated to study how city agencies currently use algorithms: [сайт]. — URL: [legistar.council.nyc.gov/LegislationDetail.aspx? ID=43137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0](http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=43137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0) (дата обращения: 21.02.2023).
4. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms //Big data & society. – 2016. – Т. 3. – №. 1. – С. 2053951715622512.
5. Ananny M., Crawford K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability //new media & society. – 2018. – Т. 20. – №. 3. – С. 973-989.
6. Fenster M. The transparency fix: Secrets, leaks, and uncontrollable government information. – Stanford University Press, 2017.
7. Grey C., Costas J. Secrecy at work: The hidden architecture of organizational life. – Stanford University Press, 2016.
8. Мартынов А. В., Бундин М. В. О правовых принципах применения искусственного интеллекта при осуществлении органами исполнительной власти контрольно-надзорной деятельности //Журнал российского права. – 2020. – №. 10. – С. 59-75.
9. Leese M. The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union //Security Dialogue. – 2014. – Т. 45. – №. 5. – С. 494-511.
10. Bundesanstalt für Finanzdienstleistungsaufsicht (2018) Big Data trifft auf künstliche Intelligenz. Herausforderungen und Implikationen für Aufsicht und Regulierung von

- Finanzdienstleistungen: [сайт]. — URL:  
[www.bafin.de/SharedDocs/Downloads/DE/dl\\_bdai\\_studie.html](http://www.bafin.de/SharedDocs/Downloads/DE/dl_bdai_studie.html) (дата обращения: 21.02.2023).
11. Tutt A. An FDA for Algorithms' (2017) //Administrative law review. – Т. 69. – С. 83.
  12. IBM. Continuous relevancy training: [сайт]. — URL:  
[console.bluemix.net/docs/services/discovery/continuous-training.html#crt](http://console.bluemix.net/docs/services/discovery/continuous-training.html#crt) (дата обращения: 21.02.2023).
  13. Hermstrüwer Y. Artificial intelligence and administrative decisions under uncertainty //Regulating Artificial Intelligence. – 2020. – С. 199-223.
  14. Lehr D., Ohm P. Playing with the data: what legal scholars should learn about machine learning //UCDL Rev. – 2017. – Т. 51. – С. 653.
  15. Воробьёва И. Б. Этические аспекты использования систем искусственного интеллекта при расследовании преступлений //Вестник Саратовской государственной юридической академии. – 2022. – №. 4 (147). – С. 162-172.
  16. Харитонова Ю. С., Савина В. С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права //Вестник Пермского университета. Юридические науки. – 2021. – №. 53. – С. 488-515.
  17. Cowgill B., Tucker C. Algorithmic bias: A counterfactual perspective //NSF Trustworthy Algorithms. – 2017.
  18. Lewis D. Counterfactuals. Harvard University Press. Cambridge, MA. – 1973.
  19. Информация Конституционно-правовая защита предпринимательства: актуальные аспекты (на основе решений Конституционного Суда Российской Федерации 2018-2020 годов) (одобрено решением Конституционного Суда РФ от 17.12.2020)
  20. SyRI legislation in breach of European Convention on Human Rights: [сайт]. — URL:  
<https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx> (дата обращения 21.02.2023)
  21. District Court of the Hague, 6 March 2020, ECLI:NL:RBDHA:2020:865: [сайт]. — URL:  
[uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878](https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878) (дата обращения 21.02.2023)
  22. MAKING O. F. A. D. PROFILING THE UNEMPLOYED IN POLAND: SOCIAL AND POLITICAL IMPLICATIONS.
  23. Koniec profilowania bezrobotnych: [сайт]. — URL:  
<https://www.prawo.pl/kadry/bezrobotni-nie-beda-profilowani-utrudnialo-to-ich-aktywizacje,394701.html> (дата обращения 21.02.2023)
  24. Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold: [сайт]. — URL: <https://spectrum.ieee.org/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold#toggle-gdpr> (дата обращения 21.02.2023)
  25. Cahoo v. SAS Analytics Inc. Nos. 18-1295/1296: [сайт]. — URL:  
<https://casetext.com/case/cahoo-v-sas-analytics-inc> (дата обращения 21.02.2023)

## **Результаты процедуры рецензирования статьи**

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

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

РЕЦЕНЗИЯ на статью на тему «Обеспечение автономности принятия решения искусственным интеллектом для целей публичных правоотношений».

Предмет исследования. Предложенная на рецензирование статья посвящена обеспечению «...автономности принятия решения искусственным интеллектом для целей публичных правоотношений». Автором выбран особый предмет исследования: предложенные вопросы исследуются с точки зрения информационного и административного права, при этом автором отмечено, что «...отдельную роль занимает вопрос позиционирования ИИ в сфере публичных правоотношений, остается системным и комплексным вопрос как на уровне теоретических исследований, так и практических решений в области понятности и прозрачности принятых ИИ решений». Изучаются НПА, зарубежная судебная практика, имеющие отношение к цели исследования. Также изучается и обобщается большой объем российской и зарубежной (на английском языке) научной литературы по заявленной проблематике, анализ и дискуссия с данными авторами-оппонентами присутствует. При этом автор отмечает: «При рассмотрении технико-правового аспекта функционирования ИИ для целей публичной власти первым аспектом является обеспечение прозрачности ИИ...».

Методология исследования. Цель исследования определена названием и содержанием работы: «...общая кодификация административного права и процесса позволяет в большинстве своем поставить деятельность служащих профильных министерств и ведомств на «алгоритмические рельсы»...», «...на уровне доктринального восприятия ИИ для целей публичных правоотношений необходимо отметить так же, что ИИ и его базисный составной элемент - машинное обучение- работают по той выборке дел, которые ранее были наработаны соответствующим должностным лицом». Они могут быть обозначены в качестве рассмотрения и разрешения отдельных проблемных аспектов, связанных с вышеназванными вопросами и использованием определенного опыта. Исходя из поставленных цели и задач, автором выбрана определенная методологическая основа исследования. Автором используется совокупность частнонаучных, специально-юридических методов познания. В частности, методы анализа и синтеза позволили обобщить подходы к предложенной тематике и повлияли на выводы автора. Наибольшую роль сыграли специально-юридические методы. В частности, автором применялись формально-юридический и сравнительно-правовой методы, которые позволили провести анализ и осуществить толкование норм актов российского законодательства, зарубежной судебной практики и сопоставить различные документы. В частности, делаются такие выводы: «Необходимо определение секторов применения ИИ, которые не влекут за собой нарушение базовых конституционных прав граждан» и др. Таким образом, выбранная автором методология в полной мере адекватна цели статьи, позволяет изучить многие аспекты темы.

Актуальность заявленной проблематики не вызывает сомнений. Данная тема является важной в мире и в России, с правовой точки зрения предлагаемая автором работа может считаться актуальной, а именно он отмечает «...нужно учитывать, что технология ИИ может быть использована не только во благо государства, но и в противовес действующим мерам, предусмотренным законодательства». И на самом деле здесь должен следовать анализ работ оппонентов, и он следует и автор показывает умение владеть материалом. Тем самым, научные изыскания в предложенной области стоит только приветствовать.

Научная новизна. Научная новизна предложенной статьи не вызывает сомнения. Она выражается в конкретных научных выводах автора. Среди них, например, такой: «... вопрос обобщения административных дел, равно как и их персонализация для целей ИИ, порождает ряд проблем, таких как проблемы эффекта отбора и сформированной

предвзятости». Как видно, указанный и иные «теоретические» выводы могут быть использованы в дальнейших исследованиях. Таким образом, материалы статьи в представленном виде могут иметь интерес для научного сообщества.

Стиль, структура, содержание. Тематика статьи соответствует специализации журнала «Административное и муниципальное право», так как посвящена обеспечению «... автономности принятия решения искусственный интеллектом для целей публичных правоотношений». В статье присутствует аналитика по научным работам оппонентов, поэтому автор отмечает, что уже ставился вопрос, близкий к данной теме и автор использует их материалы, дискутирует с оппонентами. Содержание статьи соответствует названию, так как автор рассмотрел заявленные проблемы, достиг цели своего исследования. Качество представления исследования и его результатов следует признать доработанным. Из текста статьи прямо следуют предмет, задачи, методология, результаты исследования, научная новизна. Оформление работы соответствует требованиям, предъявляемым к подобного рода работам. Существенные нарушения данных требований не обнаружены, кроме описок «как общные черты», «предсказанного преследования в виду низкого качества».

Библиография достаточно полная, содержит публикации, к которым автор обращается. Это позволяет автору правильно определить проблемы и поставить их на обсуждение. Следует высоко оценить качество представленной и использованной литературы. Присутствие научной литературы показало обоснованность выводов автора и повлияло на выводы автора. Труды приведенных авторов соответствуют теме исследования, обладают признаком достаточности, способствуют раскрытию многих аспектов темы.

Апелляция к оппонентам. Автор провел серьезный анализ текущего состояния исследуемой проблемы. Автор описывает разные точки зрения оппонентов на проблему, аргументирует более правильную по его мнению позицию, опираясь на работы оппонентов, предлагает варианты решения проблем.

Выводы, интерес читательской аудитории. Выводы являются логичными, конкретными «При внедрении полностью автономных ИИ важно учитывать сложившийся негативный международный опыт в части не прозрачных и неконституционных механизмов принятия решения ИИ, с внедрением в российском правоприменительном поле соответствующих компенсационных механизмов обеспечения прозрачности принятия решения и его верификации профильным сотрудником ФОИВа» и др. Статья в данном виде может быть интересна читательской аудитории в плане наличия в ней систематизированных позиций автора применительно к заявленным в статье вопросам. На основании изложенного, суммируя все положительные и отрицательные стороны статьи «рекомендую опубликовать».