

Информационное общество и власть

## ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ

Статья рекомендована к публикации членом редакционного совета Ю.Е. Хохловым 20.06.2021.

**Талапина Эльвира Владимировна**

*Доктор юридических наук*

*РАНХиГС при Президенте РФ, Институт прикладных экономических исследований, Центр технологий государственного управления, ведущий научный сотрудник*

*Россия, Москва*

*talapina-ev@ranepa.ru*

### Аннотация

*Технология искусственного интеллекта (ИИ) открывает перед современным обществом невероятные возможности для стремительного роста, однако она же сопряжена с существенными рисками, особенно если речь идет об использовании ИИ в сфере государственного управления. В статье приводятся преимущества использования ИИ государственным сектором, отмечаются ключевые ориентиры для использования ИИ, а также намечаются контуры правового регулирования использования ИИ в государственном управлении. Отдельно выделяются потенциальные риски, связанные с применением ИИ. Статья подготовлена на основе научно-исследовательской работы, выполняемой в рамках государственного задания РАНХиГС.*

### Ключевые слова

*Государственное управление; искусственный интеллект; правовые ограничения; риски; цифровизация; цифровые технологии*

### Введение

В современных условиях искусственный интеллект (далее – ИИ) все чаще связывают с машинным обучением. Последние достижения в области машинного обучения обусловлены целым рядом обстоятельств, как-то: совершенствование алгоритмов; увеличение финансирования; огромный рост объема данных, создаваемых и хранимых цифровыми системами; увеличение доступа к вычислительной мощности и расширение облачных вычислений. При этом существует множество подходов к определению искусственного интеллекта, однако в контексте рассматриваемой нами темы значение приобретают следующие обстоятельства: во-первых, использование статистики для поиска закономерностей в больших объемах данных, во-вторых, способность выполнять повторяющиеся задачи с данными без необходимости постоянного руководства со стороны человека.

Существует разнообразие подходов к определению искусственного интеллекта, в которых делаются разные акценты. В Российской Федерации официализировано следующее определение: искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека (Национальная стратегия развития искусственного интеллекта на период до 2030 года). В условиях глобальной цифровизации технологические инновации активно используются не только в частном, но и в государственном секторе.

---

© Талапина Э.В., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

## 1 Преимущества ИИ для государственного сектора

Идеальная память и внимательность, ориентация в бесконечных массивах данных, скорость и предвидение – такими характеристиками обладают далеко не все государственные служащие. Это создает громадный потенциал использования искусственного интеллекта. Среди преимуществ использования ИИ государственным сектором в общем плане выделяют следующие:

- обеспечение более точной информации, прогнозирования и предвидения, ведущих к улучшению результатов (например, более эффективная реализация контрольно-надзорной деятельности государственных органов);
- наличие положительного социального эффекта за счет использования ИИ в целях решения одних из самых сложных социальных проблем в мире;
- моделирование сложных систем для осуществления эксперимента с различными политическими решениями и обнаружения непредвиденных последствий перед внедрением той или иной меры;
- совершенствование процесса оказания государственных услуг (например, персонализация государственных услуг, исходя из конкретных личных обстоятельств);
- автоматизация простых ручных задач, которые освобождают кадры для более интересной работы [1].

Разумеется, в использовании ИИ нельзя видеть только сплошные плюсы. «Клиентоориентированность» технологии имеет и другую сторону медали. В частности, необходимо учитывать ряд факторов, включая этику использования ИИ и безопасность.

В юридическом плане нужно особо подчеркнуть необходимость соблюдения законов о защите персональных данных в отношении автоматизированных решений. Под последними в данном случае понимаются решения, принятые при отсутствии вмешательства со стороны человека, которые оказывают существенное юридическое или аналогичное воздействие на субъекты персональных данных. В частной сфере к таковым можно отнести онлайн-решения о предоставлении кредита или тест по подбору персонала, который использует заранее запрограммированные алгоритмы. В государственном секторе речь может идти о принятии автоматизированных управленческих решений (принимаемые в отношении конкретных лиц решения о выдаче лицензии, привлечении к ответственности). Уже оформляется целая область регулирования автоматизированных решений, в особенности в контексте законодательства о защите персональных данных. В связи с этим необходимо предоставить пользователям конкретную и легкодоступную информацию об автоматизированном процессе принятия решений, а также предусмотреть простые способы, при помощи которых будет происходить вмешательство со стороны человека в случае необходимости пересмотра и потенциального изменения автоматизированного решения [1].

## 2 Ключевые ориентиры для использования ИИ в государственном управлении

При использовании ИИ в государственном секторе также необходимо обратить особое внимание на пять ключевых моментов: 1) точность; 2) контроль со стороны человека; 3) объяснимость и прозрачность; 4) справедливость; 5) конфиденциальность и защита персональных данных, гражданские права и свободы [2]. В своей совокупности они влияют на осуществление качественного государственного управления как с процедурной, так и с содержательной стороны. Рассмотрим их подробнее.

### 2.1 Точность

Одно из очевидных преимуществ ИИ связано с тем, что он должен обеспечивать более точные результаты обработки больших данных, которые могут быть использованы при принятии решений. Использование ИИ позволяет, во-первых, рассматривать возросшее число входных переменных среди больших наборов данных систематически; во-вторых, игнорировать переменные параметры, которые не являются релевантными. Кроме того, в процессе принятия решений ИИ оценивает ошибки таким способом, который позволяет скорректировать настройки.

Инструменты ИИ должны регулярно тестироваться и «переучиваться» в целях обеспечения адекватности настроек ИИ и отражения постоянно меняющихся приоритетов правительства,

государственных стратегий, законодательства и социальных условий. В этом смысле необходим, особенно на первых порах, человеческий контроль. Например, Правительство Канады в целях обеспечения контроля качества точности результатов ввело Директиву «О консультировании при автоматизированном принятии решений» (Directive on Automated Decision-Making Consulting) [3], в рамках которой требуется, чтобы соответствующие квалифицированные эксперты проверяли автоматизированную систему принятия решений (the Automated Decision System). Согласно требованиям данной директивы необходимо наличие не менее одного квалифицированного эксперта из федерального, провинциального, территориального или муниципального государственного учреждения, одного преподавателя высшего учебного заведения и по крайней мере двух квалифицированных экспертов из Национального исследовательского совета Канады (the National Research Council of Canada), Статистического управления Канады (Statistics Canada) или Центра безопасности коммуникаций (the Communications Security Establishment). Применяемые показатели качества работы выбираются в каждом конкретном случае с учетом характера принимаемого решения, а также общих целей и ожиданий, поставленных для конкретного применения ИИ.

## 2.2 Контроль со стороны человека

Контроль со стороны человека не ограничивается только оценкой точности результатов обработки данных. Использование ИИ в процессе принятия решений предполагает ту или иную степень контроля со стороны человека. В свою очередь, степень контроля будет варьироваться в зависимости от значимости решения и от других имеющихся гарантий. В тех случаях, когда решение или сделанный в отношении индивида выбор имеет принципиальное значение (его реализация воздействует на льготы, свободу или доступ к услугам), необходимо тщательным образом рассмотреть требуемый уровень участия человека.

К примеру, в Новой Зеландии Апелляционным судом была рассмотрена апелляционная жалоба на решение о продлении приказа об осуществлении надзора (Extended Supervision Order) в отношении заключенного, который вскоре должен был быть освобожден. Первоначальное распоряжение о надзоре было подкреплено результатами сложных алгоритмических вычислений, измеряющих вероятность повторного правонарушения со стороны заключенного. Рассматривая связь между личными обстоятельствами заявителя и полученными вероятностными результатами, Суд высказал следующее замечание: «Очевидно, что факторы, возникшие после освобождения, должны быть учтены при оценке продления приказа об осуществлении надзора. Например, если бы после освобождения заявитель получил тетраплегию в результате несчастного случая, это, по-видимому, исключило бы вероятность повторного правонарушения с его стороны и, несомненно, отрицало бы любые неблагоприятные выводы, которые в противном случае могли бы иметь место для вероятностных оценок.» [1].

## 2.3 Транспарентность

Одной из наиболее фундаментальных проблем, связанных с использованием ИИ государственным сектором, является потенциальная возможность отсутствия открытости используемых алгоритмов. Этот момент становится еще более значительным в свете декларируемой многими государствами транспарентности государственного управления.

На техническом уровне ИИ можно представить в качестве процесса, подобного «черному ящику». Черный ящик – это описание, применяемое к некоторым системам глубокого обучения, которые принимают входной сигнал и обеспечивают выходной, однако происходящие между этими процессами вычисления нелегко интерпретировать людям. Системы ИИ типа «черный ящик» принимают решения экспериментально или «интуитивно», не давая ответ на вопрос, почему было принято то или иное решение.

Проблема заключается в том, что содержательные объяснения алгоритма возможны, но они могут быть недоступны. Например, право интеллектуальной собственности может препятствовать раскрытию проприетарного кода или доступу к обучающим данным, поэтому даже если и можно было бы понять, как работает алгоритм, полные вычисления могут оказаться невозможными по экономическим, юридическим или политическим причинам.

В связи с этим Руководство Правительства Канады «Digital Playbook Guide on automated decisions» рекомендует сделать доступным для общественности весь исходный код, используемый для их автоматизированных систем принятия решений, и требует, чтобы заинтересованным лицам

было предоставлено подробное объяснение, включая переменные параметры в решении, а также само решение [4].

#### **2.4 Справедливость**

Нужно иметь в виду, что инструменты ИИ программируются людьми, а также то, что предвзятость данных (в отличие от предвзятости человека-оператора) создает значительную проблему для эффективного использования ИИ.

Даже лучшие инструменты ИИ могут увековечить историческое неравенство, если предвзятость в данных не определена и не учтена, а приоритеты системы не будут согласованы с ожиданиями справедливости. Другими словами, там, где инструмент ИИ применяется к неравным условиям, и это не было изначально допущением, могут возникнуть искаженные результаты. Это особенно актуально при рассмотрении уязвимых или находящихся в неблагоприятном положении членов общества, таких как коренные народы, национальные меньшинства.

Данные риски могут быть несколько нивелированы, если ИИ используется для оказания поддержки в принятии решений человеком, а не для полной замены человека. Кроме того, необходим процесс, гарантирующий, что нерелевантные и/или несправедливые, причиняющие вред характеристики, будут специально исключаться из информации, предоставляемой для программирования и обучения ИИ.

#### **2.5 Конфиденциальность и защита персональных данных**

Законодательные режимы различаются в зависимости от юрисдикции, но основополагающие вопросы защиты персональных данных и конфиденциальности информации, а также связанные с ними гражданские права и свободы, которые затрагиваются использованием ИИ, относительно универсальны.

Данные, используемые для «питания» и обучения систем искусственного интеллекта, часто представляют собой комбинацию неперсональных и персональных данных. Основанием получения государственными учреждениями такой информации, как правило, выступает возможность ее использования в разведывательных целях; в той мере, в которой она является персональной, она будет анонимизирована. По иронии судьбы, та же самая технология, которая позволяет ИИ анализировать информацию, теперь представляет собой риск, связанный со способностью ИИ повторно идентифицировать ранее анонимизированные данные, тем самым вступая в противоречие с основанием, в соответствии с которым была собрана информация. Риски повторной идентификации повышаются также при межведомственном обмене информацией.

### **3 Контуры правового регулирования использования ИИ в государственном управлении**

Такая специфическая сфера, как государственное управление, требует продуманного правового регулирования. Использование искусственного интеллекта не должно нарушать сложившийся баланс между публичными интересами и правами граждан. Поэтому внедрение ИИ необходимо конструировать с учетом того потенциального вреда, который могут принести системы ИИ, и содержать механизмы его предотвращения/минимизации. Рассмотрим, о каких рисках идет речь [5].

#### **3.1. Предвзятость и дискриминация**

Поскольку технологии, основанные на данных, черпают свое понимание из существующих общественных отношений, которые они анализируют, постольку данные технологии могут воспроизводить, усиливать и расширять уже существующие в обществе паттерны маргинализации, неравенства и дискриминации [6]. Кроме того, многие функции, метрики и аналитические структуры моделей, обеспечивающие интеллектуальный анализ данных, выбираются их разработчиками, тем самым эти технологии потенциально могут воспроизводить предрассудки и предубеждения их разработчиков.

Наконец, выборки данных, используемые для обучения и тестирования алгоритмических систем, часто могут быть недостаточно репрезентативными для тех популяций, в отношении которых они делают выводы. Это создает реальные возможности предвзятых и дискриминационных результатов, поскольку данные, поступающие в систему, изначально дефектны. Нужно иметь в виду, что дискриминация посредством алгоритма имеет систематический характер, что на деле намного вреднее, чем разовый людской субъективизм [7].

Для правового регулирования, в основе которого в демократическом государстве лежит принцип равенства, такой риск может стать критичным.

### **3.2. Отказ от автономии личности**

Когда граждане подчиняются решениям, прогнозам или классификациям, производимым системами ИИ, могут возникнуть ситуации, при которых они не смогут возложить прямую ответственность на тех лиц, которые отвечают за эти результаты. Системы ИИ автоматизируют те когнитивные функции, которые ранее относились исключительно к компетенции человека, ответственного за осуществляемую им деятельность. Сказанное может усложнить определение ответственности в алгоритмически генерируемых результатах, поскольку сложный и распределенный характер процессов проектирования, производства и внедрения систем ИИ может вызвать затруднение в процессе определения ответственных сторон. В случае причинения вреда или наступления иных негативных последствий проблема недостаточного привлечения виновных к ответственности может нанести ущерб автономии личности и привести к нарушению прав пострадавших лиц.

### **3.3. Нетранспарентные, неподдающиеся объяснению или неоправданные результаты**

Многие модели машинного обучения генерируют результаты, оперируя высокомерными корреляциями, которые находятся за пределами интерпретационных возможностей человеческого мышления. В этих случаях обоснование алгоритмически полученных результатов, которые непосредственно влияют на субъекты принятия решений, остается нетранспарентным для этих же субъектов. В то время как в одних случаях отсутствие объяснения того, как работает алгоритм, может быть приемлемым, в других это представляет собой проблему, например, когда обрабатываемые данные могут скрывать следы дискриминации, предвзятости, неравенства или несправедливости.

### **3.4. Вторжение в частную жизнь**

Угрозы конфиденциальности персональных данных, исходящие от систем ИИ, существуют как в результате их проектирования и разработки, так и в процессе их развертывания. Поскольку проекты ИИ связаны со структурированием и обработкой данных, постольку развитие технологий ИИ часто будет включать использование персональных данных, причем иногда без согласия субъекта данных.

Системы ИИ, которые таргетируют, осуществляют выборку по определенному критерию из массива данных и без ведома и согласия субъектов данных «подталкивают» их к определенному выбору, в некоторых случаях могут оказывать существенное воздействие на частную жизнь субъектов данных в том смысле, что трансформирующая природа технологий ставит под сомнение способность индивидуума к управлению своей собственной жизнью. Подобного рода вторжение в частную жизнь способно, следовательно, нанести ущерб основному праву человека преследовать свои жизненные цели и ориентиры, свободные от влияния извне.

### **3.5. Изоляция и дезинтеграция социальных связей**

Способность систем ИИ специально отбирать индивидуальный опыт и персонализировать цифровые услуги обещает не только значительно улучшить жизнь потребителей и повысить эффективность предоставления услуг, но и сопряжена с потенциальными рисками. Избыточная автоматизация, например, может уменьшить потребность людей во взаимодействии друг с другом, в свою очередь, сопровождаемая поддержкой алгоритмов гиперперсонализация может поляризовать социальные отношения посредством ограничения нашего воздействия на мировоззрение, отличное от нашего. То, как быстро люди переключаются на онлайн-общение и утрачивают социальные привычки, продемонстрировала пандемия коронавируса. Между тем, хорошо организованные и сплоченные общества строятся на доверительных отношениях, сопереживании и взаимопонимании. По мере того, как технологии ИИ становятся все более распространенными, важно, чтобы эти отношения сохранялись.

### **3.6. Неадекватные, небезопасные или некачественные результаты**

Безответственное управление данными, небрежное проектирование и осуществление производственных процессов в сочетании с сомнительными методами развертывания могут, каждый по-своему, привести к внедрению и распространению систем ИИ, которые дают

ненадежные, небезопасные или некачественные результаты. Кроме того, это может подорвать доверие общественности к ответственному использованию социально полезных технологий ИИ [5].

Такие серьезные риски можно преодолеть только адекватным и гибким законодательством. Очевидно, что использование ИИ в государственном управлении возможно как на уровне правотворчества (например, в законопроектной деятельности [8,9]), так и на уровне правоприменения (установление лжи в процессе таможенного контроля), а также вплоть до создания персонализированного законодательства. Утверждается, что введение ИИ повлечет появление нового права человека - на прозрачность алгоритмов [10]. Все изложенные обстоятельства очерчивают контуры будущего правового регулирования использования ИИ в государственном управлении.

## Заключение

Привлекательность внедрения ИИ в государственное управление очевидна. Однако специфика этой отрасли требует особо внимательного отношения к правам граждан, к возможностям оперативного реагирования на их нарушения, к прозрачности государственной деятельности. У людей должна остаться возможность получения государственных услуг в привычном формате, путем обращения к государственному служащему - человеку. Иными словами, нельзя допустить революционного изменения юридической культуры. Оптимисты утверждают, что юристы всегда будут иметь важное значение для привития правовых ценностей и совершенствования правовой культуры - так будет и сейчас, и в будущем, когда мы будем жить в мире, где работают как юристы с ИИ, так и люди [11]. Только построение правового регулирования с учетом традиционных этических ценностей и завоеваний доцифрового права создаст комфортную среду взаимодействия ИИ с людьми на одном из самых ответственных участков - в государственном управлении.

## Литература

1. A guide to using artificial intelligence in the public sector // URL: <https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector> (дата обращения: 28.02.2021).
2. The use of Artificial Intelligence to Combat Public Sector Fraud. Professional Guidance // International Public Sector Fraud Forum. 2020. 40 p.
3. Directive on Automated Decision-Making // URL: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> (дата обращения: 17.03.2021).
4. Government of Canada Digital Playbook (draft) // URL: <https://canada-ca.github.io/digital-playbook-guide-numerique/en/overview.html> (дата обращения: 17.03.2021).
5. Leslie D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector // The Alan Turing Institute, 2019. 97 p.
6. Hacker Ph. The Ambivalence of Algorithms. Gauging the Legitimacy of Personalized Law // Bakhoun Mor, Conde Gallego Beatriz, Mackenrodt Mark-Oliver, Surblytė-Namavičienė Gintarė (Eds.). Personal Data in Competition, Consumer Protection and Intellectual Property Law. Towards a Holistic Approach? Springer, 2018. Pp. 85-117.
7. Faisal K., Indrė Žl, Toon C. Quantifying explainable discrimination and removing illegal discrimination in automated decision making // Knowledge and Information Systems, 2013. Pp. 613-644.
8. Соколова А.А. Искусственный интеллект в юриспруденции: риски внедрения // Юридическая техника, 2019. № 13. С. 350-356.
9. Wickers Th. Peut-on coder le droit? // Cahiers de droit de l'entreprise, 2019. № 4. P.30.
10. Pour une intelligence artificielle maîtrisée, utile et démystifiée. Rapport de L'office Parlementaire d'évaluation des choix Scientifiques et Technologiques. URL: <http://www.assemblee-nationale.fr/14/pdf/rap-off/i4594-tII.pdf> (дата обращения: 15.03.2021).
11. Dervanović D. I. Inhuman Lawyer: Developing Artificial Intelligence in the Legal Profession // Corrales M., Fenwick M., Forgó N. (Eds.) Robotics, AI and the Future of Law. Springer, 2018. Pp. 209-234.

# THE USE OF ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATION

Talapina, Elvira V.

*Doctor of legal sciences*

*RANEPa, AERI, Public Management Technologies Center, leading researcher*

*Moscow, Russian Federation*

*talapina-ev@ranepa.ru*

## Abstract

*In modern society artificial intelligence (AI) technology opens incredible opportunities for rapid growth, although it also provides significant risks, especially when it comes to the use of AI in public administration. The article presents the advantages of the using artificial intelligence in the public sector, highlights the key guidelines of AI using, and outlines the framework of the legal regulation in the context of AI using in public administration. The potential risks associated with AI using are highlighted particularly. The article was written on the basis of the RANEPa state assignment research program.*

## Keywords

*Public administration; artificial intelligence; legal restrictions; digitalization; digital technologies*

## References

1. A guide to using artificial intelligence in the public sector // URL: <https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector> (accessed: 28.02.2021).
2. The use of Artificial Intelligence to Combat Public Sector Fraud. Professional Guidance // International Public Sector Fraud Forum. 2020. 40 p.
3. Directive on Automated Decision-Making // URL: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> (accessed: 17.03.2021).
4. Government of Canada Digital Playbook (draft) // URL: <https://canada-ca.github.io/digital-playbook-guide-numerique/en/overview.html> (accessed: 17.03.2021).
5. Leslie D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector // The Alan Turing Institute, 2019. 97 p.
6. Hacker Ph. The Ambivalence of Algorithms. Gauging the Legitimacy of Personalized Law // Bakhoun Mor, Conde Gallego Beatriz, Mackenrodt Mark-Oliver, Surblytė-Namavičienė Gintarė (Eds.). Personal Data in Competition, Consumer Protection and Intellectual Property Law. Towards a Holistic Approach? Springer, 2018. Pp. 85-117.
7. Faisal K., Indrė Žl, Toon C. Quantifying explainable discrimination and removing illegal discrimination in automated decision making // Knowledge and Information Systems, 2013. Pp. 613-644.
8. Sokolova A.A. Iskusstvennyj intellekt v yurisprudencii: riski vnedreniya // Yuridicheskaya tekhnika. 2019. № 13. S. 350-356.
9. Wickers Th. Peut-on coder le droit? // Cahiers de droit de l'entreprise. 2019. № 4. P.30.
10. Pour une intelligence artificielle maîtrisée, utile et démystifiée. Rapport de L'office Parlementaire d'évaluation des choix Scientifiques et Technologiques. URL: <http://www.assemblee-nationale.fr/14/pdf/rap-off/i4594-tII.pdf> (accessed: 15.03.2021).
11. Dervanović D. I. Inhuman Lawyer: Developing Artificial Intelligence in the Legal Profession // Corrales M., Fenwick M., Forgó N. (Eds.) Robotics, AI and the Future of Law. Springer, 2018. Pp. 209-234.