EDN CLJCAK

ИИ-ориентированные государственные сервисы: таксономия ответственности и суверенный искусственный интеллект

Носиков А. А.

Санкт-Петербургский государственный университет, Российская Федерация; a.nosikov@spbu.ru

РЕФЕРАТ

Настоящее исследование анализирует институциональные и технологические вызовы интеграции систем искусственного интеллекта (ИИ) в публичное администрирование и государственные сервисы, фокусируясь на классификации ролей алгоритмов в процессах принятия решений, балансе интересов в сотрудничестве с коммерческими поставщиками ИИ-решений и инфраструктуры, а также обеспечении национальной технологической автономии. Применен качественный междисциплинарный подход, сочетающий нормативно-правовой анализ, тематический анализ эмпирических кейсов из практики различных стран и теоретический синтез. Данные собраны из официальных источников, рецензируемых научных публикаций и новостных источников с использованием метода снежного кома для отбора кейсов, а кодирование проводилось итеративно. В результате разработана оригинальная авторская шестиуровневая пирамидальная модель распределения ответственности в зависимости от степени автономии алгоритмов ИИ в цепочке принятия решений: от полной делегации («ИИ-Капитан») через предложение готового решения с утверждением человеком («ИИ-Штурман»), набор конфигураций («ИИ-Советник»), анализ среды с сигнализацией триггеров («ИИ-Наблюдатель»), выполнение трудозатратных задач с ревизией оператором («ИИ-Рабочие руки») до рутинной поддержки без решений («ИИ-Рутинный помощник»). Модель наложена на градации рисков (высокий, ограниченный, минимальный) для оценки последствий ошибок. Выявлена дилемма государственно-частного партнерства, обеспечивающего доступ к инновациям, но усиливающего зависимость и уязвимости. Также обоснована роль суверенного ИИ как стратегии снижения этих рисков. Для эффективной интеграции алгоритмов в государственные сервисы рекомендуется внедрение обязательной классификации систем ИИ по уровням автономии и критичности, где шестиуровневая таксономия обеспечивает дифференцированный подход к распределению ответственности, минимизируя институциональные пробелы и риски предвзятости.

Ключевые слова: искусственный интеллект, ИИ-ориентированные государственные сервисы, публичное администрирование, суверенный ИИ, институциональные риски, цифровая трансформация.

Для цитирования: *Носиков А. А.* ИИ-ориентированные государственные сервисы: таксономия ответственности и суверенный искусственный интеллект // Управленческое консультирование. 2025. № 5. С. 65–76. EDN CLJCAK

AI-Driven Public Services: A Taxonomy of Accountability and Sovereign Artificial Intelligence (AI)

Andrey A. Nosikov

Saint Petersburg State University, Russian Federation; a.nosikov@spbu.ru

ABSTRACT

This study examines the institutional and technological challenges of integrating artificial intelligence (AI) systems into public administration and governmental services, focusing on the taxonomy of algorithmic roles in decision-making, the balance of interests in cooperation with commercial AI providers and infrastructure actors, and the safeguarding of national technological sovereignty. A qualitative interdisciplinary approach is applied, combining regulatory and legal analysis, thematic examination of empirical cases across different countries, and theoretical

synthesis. Data were collected from official documents, peer-reviewed publications, and news sources, using snowball sampling for case selection and iterative coding for analytical categorization.

The research develops a six-tier pyramidal model of accountability distribution according to the degree of algorithmic autonomy in decision-making chains: from full delegation («AI as Captain»), provision of ready-made solutions for human approval («AI as Navigator»), configuration of option sets («AI as Adviser»), environmental analysis with trigger signaling («AI as Observer»), execution of labor-intensive tasks under operator supervision («AI as Workforce»), to routine operational support without decision-making capacity («AI as Routine Assistant»). The model is mapped against risk gradations (high, limited, minimal) to assess error consequences.

The findings reveal the dilemma of public-private partnerships, which facilitate access to innovation but simultaneously reinforce dependence and systemic vulnerabilities. The study also substantiates the role of sovereign AI as a strategic response to these risks. For effective integration of AI into governmental services, it recommends mandatory classification of systems by autonomy and criticality levels. The proposed six-level taxonomy enables a differentiated approach to accountability allocation, reducing institutional gaps and risks of bias, while enhancing resilience and strategic security.

Keywords: artificial intelligence, AI-Driven public services, public administration, sovereign AI, institutional risks, digital transformation.

For citation: Nosikov A. A. AI-Driven Public Services: A Taxonomy of Accountability and Sovereign Artificial Intelligence (AI) // Administrative Consulting. 2025. N 5. P. 65–76. EDN CLJCAK

Введение

В последние годы государства активизировали внедрение сервисов и инфраструктур на основе искусственного интеллекта (ИИ). Трансформация публичного администрирования охватывает оптимизацию процессов и автоматизацию рутинных действий [8], поддержку стратегических решений в общественной безопасности, здравоохранении и управлении критической инфраструктурой [12]. Она создает значительные институциональные и технологические возможности [21], но одновременно порождает новые риски и дилеммы, связанные с ответственностью, легитимностью и суверенитетом цифрового управления [10].

Актуальность исследования определяется несколькими факторами. Во-первых, экспансия ИИ в публичный сектор меняет распределение ролей между государством, гражданами и коммерческими поставщиками технологий [18], делая критически важными вопросы о том, кто и на каких основаниях принимает решения для обеспечения правовой определенности и защиты прав человека. Во-вторых, усиление зависимости государственных функций от коммерческих платформ и облачной инфраструктуры ставит проблему баланса между эффективностью и национальной технологической безопасностью [28]. В-третьих, разнообразие уровней автономности ИИ в цепочке принятия решений требует дифференцированного подхода к регуляции, распределению ответственности и оценке критичности последствий алгоритмических ошибок [2]. Эти факторы подчеркивают прикладную значимость исследования внедрения ИИ-ориентированных государственных сервисов.

Проблематика исследования многослойна. В центре находится таксономия ответственности, формализующая роль ИИ в принятии решений и определяющая институциональные и юридические субъекты ответственности при различных уровнях автономности алгоритмов [26]. Анализируется также дилемма государственно-частного партнерства: сотрудничество с технологическими гигантами ускоряет внедрение и повышает качество ИИ-сервисов, но увеличивает системную уязвимость, риск утечки данных [16] и зависимость госинститутов от внешних акторов [7]. Третий элемент — концепция суверенного ИИ, рассматриваемая как стратеги-

ческая реакция на платформенную зависимость и объект политики национальной безопасности, создающий собственные экономические, этические и нормативные сложности [1].

Исследовательский дизайн и методы

Настоящая статья опирается на междисциплинарный подход, объединяющий нормативно-правовой анализ, изучение кейсов внедрения ИИ-ориентированных государственных сервисов и систематизацию концепций ответственности и суверенитета в цифровую эпоху. Исследование использует качественный дизайн, акцентируя внимание на таксономии ответственности ИИ, дилеммах государственно-частного партнерства и концепции суверенного ИИ. Такой подход релевантен в условиях новизны и контекстной вариативности применения ИИ, а также различий регуляторных рамок в разных странах. Интеграция анализа научных источников [29], эмпирических кейсов и теоретического синтеза обеспечивает структурированное понимание вызовов.

Данные собирались из первичных (правительственные отчеты, нормативные документы, материалы поставщиков ИИ) и вторичных источников (рецензируемые статьи, публикации СМИ), что позволило учесть глобальные практики внедрения. Поиск научных источников осуществлялся в Google Scholar, Scopus, Web of Science, выборка формировалась методом «снежного кома». Кейсы отбирались на основе анализа инициатив Китая, США, Сингапура, Испании, Великобритании, Албании, Японии и прочих стран.

Разработка таксономии (рис. 1) проводилась посредством метода тематического анализа с последующей систематизацией категорий [3]. Кодирование данных осуществлялось итеративно: первичные коды фиксировали роли ИИ в принятии решений («автономная», «консультативная» и иные), затем объединялись в тематические кластеры, образовавшие шестиуровневую пирамиду ответственности ИИ. Критичность рисков соотносилась с этой структурой на основе дедуктивного кодирования. При составлении аналитической матрицы (см. табл. 1) использовались системный и структурный подходы.

Рис. 1. Таксономия классов ИИ по степени автономии в цепочке принятия решений
Fig. 1. Taxonomy of AI classes based on the degree of autonomy in the decision-making chain

Источник: Составлено автором.

Таксономия ответственности ИИ в цепочке принятия решений ИИ-ориентированных государственных сервисов

ИИ-ориентированные государственные сервисы можно классифицировать по уровню ответственности ИИ в цепочке принятия решений. Эта классификация имеет пирамидальную структуру: на вершине — полная автономия ИИ, в основании — отсутствие делегирования решений ИИ (см. рис. 1). Идея пирамидальной метафоры обусловлена тем, что она наглядно отражает нисходящий континуум автономности и ответственности: вверху оказываются системы с наибольшей автономией (и, соответственно, наименьшим участием человека), в основании находятся самые ограниченные автоматизированные подсистемы. Такая структура подчеркивает, что с ростом автономии алгоритма растут и критичность принятого им решения, и связанные с этим риски. Метафора пирамиды полезна еще и тем, что визуально сигнализирует о том, что число применений и масштаб применения ИИ обычно увеличиваются при переходе к низшим, менее автономным уровням, тогда как уровни вершины специфичны и реализуются реже.

Так, на высшем уровне процесс принятия решений полностью делегируется ИИ. Такой автономный класс в пирамиде ИИ-ориентированных сервисов можно метафорически обозначить как «ИИ-Капитан». Здесь возникает проблема институционализации ответственности [4], поскольку ИИ не обладает юридической субъектностью ни в одной системе права. Полное делегирование решений ИИ приводит к отсутствию ответственного субъекта в случае ошибки [23]. Примером служит проект в Китае, реализованный совместно с Alibaba. Так, в Ханчжоу внедрена система управления дорожным движением на базе ИИ, которая в реальном времени обрабатывает данные с камер и датчиков, самостоятельно оптимизирует работу светофоров, обнаруживает аварии и прогнозирует заторы. Операторы присутствуют лишь для внештатных ситуаций, тогда как основной массив решений принимается ИИ совершенно автономно[1].

Второй по уровню автономности класс — «ИИ-Штурман», который формирует готовое решение, но его утверждение осуществляется гражданином, должностным лицом или коллегиальным органом. Здесь появляется субъект ответственности, принимающий или отклоняющий предложение ИИ. Примером служит Gotham от Palantir[2] — операционная система, предлагающая сценарии реагирования на тактические и стратегические военные задачи, оставляя окончательное решение за командованием. Если говорить о востребованности таких решений, то в 2023 г. Департамент обороны США заключил контракт с Palantir на 250 млн долларов для партнерства в сфере оборонных ИИ-сервисов[3].

Следующий класс — «ИИ-Советник» — предполагает, что ИИ предлагает набор конфигураций возможных решений и переменных, однако выбор, корректировка и утверждение остаются за гражданином, должностным лицом или иной институцией. К данному классу относятся чат-боты. Так, в Сингапуре внедрены виртуальные ассистенты на базе ИИ для оказания госуслуг и обработки запросов. Примером является чат-бот Ask Jamie, функционирующий на официальных правительственных

[1] Alibaba's 'City Brain' is slashing congestion in its hometown [Электронный ресурс] // CNN: Cable News Network. URL: https://edition.cnn.com/2019/01/15/tech/alibaba-city-brain-hangzhou/index.html (дата обращения: 21.08.2025).

[2] Gotham. The Operating System for Global Decision Making [Электронный ресурс] // Palantir: официальный сайт. URL: https://www.palantir.com/platforms/gotham/ (дата обращения: 22.08.2024).

[3] Palantir Wins $250 Million AI Deal With US Defense Department [Электронный ресурс] // BNN Bloomberg. URL: https://www.bnnbloomberg.ca/palantir-wins-250-million-ai-deal-with-us-defense-department-1.1977297 (дата обращения: 23.08.2025).

сайтах и помогающий гражданам искать информацию и совершать транзакции с госведомствами[4].

Четвертый класс — «ИИ-Наблюдатель», самостоятельно анализирует среду и данные, сигнализируя о наступлении триггеров, тогда как контроль и реагирование остаются за должностным лицом, коллегиальным органом или иной институцией. Так, городской совет Барселоны применяет ИИ для предиктивного обслуживания объектов инфраструктуры, включая уличное освещение и канализацию. Алгоритмы машинного обучения анализируют данные с датчиков, прогнозируют необходимость ремонта и позволяют переходить к упреждающим работам, снижая расходы [19].

Пятый класс — это «ИИ-Рабочие руки». Он выполняет сложные задачи, такие как анализ больших массивов данных, не неся критической ответственности за окончательные решения. Результаты проверяются должностным лицом, коллегиальным органом или иной институцией. Например, управление по визам и иммиграции Великобритании (UKVI) использует ИИ для анализа документов к заявлениям на визы. Система автоматически проверяет подлинность и действительность документов, оптимизируя процесс и снижая нагрузку на сотрудников. Несмотря на ограниченную автономность, алгоритм UKVI уже подвергался обвинениям в предвзятости[5].

Нижестоящий класс — «ИИ-Рутинный помощник», выполняющий рутинные задачи без принятия решений. Так, правительство Албании использует ChatGPT для перевода тысяч страниц правовых актов ЕС на албанский язык с последующей интеграцией в национальное законодательство[6]. В этом случае ИИ выполняет рутинную работу, далекую от уровня принятия решений, а сам процесс предполагает ревизию и контроль результатов деятельности ИИ на всех этапах процедуры, а окончательная интеграция результатов остается за институциональными механизмами.

Вместе с этим важно обсудить в контексте представленной выше таксономии ответственности ИИ компонент уровня критичности ошибки ИИ и рисков в цепочке принятия решений. Например, если чат-бот уровня «ИИ-Советник» предложит гражданину некорректное решение в части навигации по сайту государственных услуг, последствия будут куда менее критичными, нежели решение, предложенное в области обороны или силовых ведомств. Таким образом, все области внедрения ИИ-ориентированных государственных сервисов можно ранжировать по уровню критичности ошибки. Безусловно, для решения этой задачи требуется много раундов консультаций с экспертами в каждой из областей внедрения ИИ в государственные сервисы, однако на предпроектном уровне можно пользоваться классификацией, предложенной Европейской Комиссией в AI Act[7]. Так, к области «высокий риск» относятся технологии ИИ, применяемые в области критически важной инфраструктуры или принятия решений, когда под угрозу могут быть поставлены жизнь и здоровье граждан, а также их права, свободы и возможности. К области «ограниченный риск» относятся вопросы, связанные с отсутствием прозрачности в деятельности

[4] Get to know the GovTech team behind Ask Jamie, the government chatbot [Электронный ресурс] // Tech.gov.sg: Government Technology Agency of Singapore. Официальный сайт. URL: https://www.tech.gov.sg/technews/govtech-team-behind-ask-jamie-government-chatbot (дата обращения: 24.08.2025).

[5] AI system for granting UK visas is biased, rights groups claim [Электронный ресурс] // The Guardian. Guardian News & Media Limited. 2019. 29 октября. URL: https://www.theguardian.com/uk-news/2019/oct/29/ai-system-for-granting-uk-visas-is-biased-rights-groups-claim (дата обращения: 24.08.2025).

[6] Albania to speed up EU accession using ChatGPT [Электронный ресурс] // EURACTIV. URL: https://www.euractiv.com/section/politics/news/albania-to-speed-up-eu-accession-using-chatgpt/ (дата обращения: 24.08.2025).

[7] AI Act [Электронный ресурс] // European Commission: официальный сайт. Shaping Europe's digital future. URL: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (дата обращения: 24.08.2025).

ИИ. Например, при использовании систем искусственного интеллекта, таких как чат-боты, люди должны знать, что они взаимодействуют с машиной, чтобы самостоятельно верифицировать окончательное решение. И наконец, «минимальный риск или его отсутствие» относится к ИИ с потенциально наименьшей степенью угрозы для благополучия граждан, например, видеоигры с интегрированным ИИ или спам-фильтры с использованием ИИ.

Дилемма государственно-частного партнерства и проблема суверенного ИИ

Важным аспектом является уровень государственно-частного партнерства в предоставлении ИИ-ориентированных государственных сервисов [17]. Так, департамент полиции Лос-Анджелеса (LAPD) использует PredPOL — программное обеспечение для прогнозирующей полицейской деятельности, анализирующее отчеты о прошлых преступлениях, местоположения и время для предсказания вероятных преступлений, что повышает эффективность распределения ресурсов и общественную безопасность[8]. PredPOL разработан коммерческой компанией SoundThinking, Inc., предоставляющей доступ к своим решениям государственным и частным субъектам[9]. Этот кейс, аналогично партнерству Palantir с министерством обороны США, иллюстрирует эффективное государственно-частное сотрудничество в сфере внедрения ИИ-ориентированных сервисов.

В свою очередь, существует иной подход, предусматривающий становление суверенных ИИ [22]. Суверенный искусственный интеллект — это концепция, обозначающая искусственный интеллект, находящийся в собственности и под суверенным контролем отдельного государства [9]. Данная парадигма имеет значение для национальной безопасности, системы государственного управления и международных отношений [6]. Концепция суверенного ИИ приобрела значительную актуальность в контексте формирования национальных стратегий развития и применения технологий искусственного интеллекта [11].

В условиях, когда государства стремятся использовать потенциал искусственного интеллекта для обеспечения экономического, оборонного и общественного развития [1], создание суверенных возможностей в области ИИ стало ключевым элементом стратегического планирования и выработки политик в этой области [24]. Взаимосвязь искусственного интеллекта и национального суверенитета также стимулирует дискуссии, касающиеся этических принципов [25], вопросов управления и регулирования ИИ в контексте национальной обороны и безопасности [15]. Внедрение технологий искусственного интеллекта в военной сфере и модернизация вооружений, включая системы высокоточного огневого поражения [13], подчеркивают критическую важность суверенного ИИ для конфигурации будущих парадигм безопасности.

Компонент суверенизации ИИ и государственно-частного партнерства в этой области сложно поддается дифференцированию, поскольку предполагает в себе дилемму. С одной стороны, при эффективном партнерстве с частными компаниями государство получит доступ к множеству конкурентных технологических решений [27]. Сегодня крупные технологические компании являются локомотивом отрасли [14], поэтому на коротком отрезке времени стратегия кооперации государств и корпоративных технологических гигантов выглядит наиболее результативной [20],

[8] Maximize Limited Patrol & Analyst Resources for Highest Impact [Электронный ресурс] // SoundThinking: официальный сайт. URL: https://www.soundthinking.com/law-enforcement/resource-deployment-resourcerouter/ (дата обращения: 25.08.2025).

[9] About SoundThinking [Электронный ресурс] // SoundThinking: официальный сайт. URL: https://www.soundthinking.com/company/ (дата обращения: 25.08.2025).

обеспечивая доступ к передовым программно-аппаратным комплексам для развертывания ИИ-ориентированых государственных сервисов и решений [30].

Важен также компонент бюджетирования. Так, Microsoft совместно с OpenAI планирует к 2028 г. развернуть центр обработки данных с суперкомпьютером Stargate, сметная стоимость которого может достигать 100 млрд долларов[10]. Если бы подобные инвестиции финансировались государством за счет налогоплательщиков, они могли бы вызвать общественные споры. В рамках государственно-частного партнерства расходы и риски ложатся на частных партнеров.

Поэтому многие правительства успешно кооперируются с технологическими компаниями для создания ИИ-ориентированных государственных сервисов. Так, правительство Японии заключило стратегическое партнерство с NEC для разработки систем реагирования на стихийные бедствия класса «ИИ-Советник»[11]. Эти системы используют спутниковое зондирование, анализ данных из социальных сетей и алгоритмы машинного обучения для оценки ущерба, координации спасательных операций и обеспечения ситуационной осведомленности в режиме реального времени[12].

Рынок готовых решений для ИИ-ориентированных сервисов широк. Так, IBM предлагает продукты для комплексной экосистемы госуслуг[13]: IBM MaS360 для защиты устройств и данных, IBM Maximo для оптимизации управления активами, чат-бот IBM Watsonx для коммуникации с гражданами 24/7, IBM Flashsystem для модернизации инфраструктуры и управления большими данными и IBM watsonx. governance как интегративная платформа для управления всеми ИИ-сервисами, включая модели от сторонних поставщиков[14].

С другой стороны, приведенные выше аргументы суверенизации ИИ в критически важных областях действительно остро поднимают вопросы о стратегической безопасности и развитии государств. Кроме того, такие риски, как банкротство компании — поставщика ИИ решений, утечка данных, инфраструктурная зависимость, риск коммерческого шпионажа, также склоняют к стимулированию разработки государствами суверенных ИИ. По этому поводу основатель технологической компании Oracle Ларри Эллисон в своем недавнем интервью заявил: «Все правительства, практически без исключений, будут стремиться к созданию суверенного искусственного интеллекта в облаке»[15].

Обсуждение

Пирамидальную таксономию, представленную выше (см. рис. 1), можно соотнести также с аналитической матрицей характеристик, включающей такие параметры, как уровни рисков и «критичность компонента суверенного ИИ» (см. табл. 1).

[10] Microsoft, OpenAI plan $100 billion data-center project, media report says [Электронный ресурс] // Reuters. URL: https://www.reuters.com/technology/microsoft-openai-planning-100-billion-data-center-project-information-reports-2024-03-29/ (дата обращения: 24.08.2025).

[11] NEC develops technology for disaster damage assessment using a Large Language Model (LLM) and image analysis [Электронный ресурс] // NEC: официальный сайт. URL: https://www.nec.com/en/press/202308/global_20230825_02.html (дата обращения: 24.08.2025).

[12] Using large language models and image analysis for disaster damage assessment [Электронный ресурс] // Tech Wire Asia. URL: https://techwireasia.com/08/2023/how-to-use-large-language-models-and-image-analysis-for-disaster-damage-assessment/ (дата обращения: 24.08.2025).

[13] Manage and protect government employees' devices, apps and data [Электронный ресурс] // IBM: официальный сайт. URL: https://www.ibm.com/products/maas360/government (дата обращения: 25.08.2025).

[14] watsonx.governance [Электронный ресурс] // IBM: официальный сайт. URL: https://www.ibm.com/products/watsonx-governance (дата обращения: 25.08.2025).

[15] Oracle's Larry Ellison thinks every government will want to build a 'sovereign' AI cloud in the future [Электронный ресурс] // CNBC. 2024. 7 апреля. URL: https://www.cnbc.com/2024/04/07/oracle-chatgpt-and-the-sovereign-cloud-nations-will-seek-in-future.html (дата обращения: 24.08.2025).

Таким образом, эмпирические данные и концептуальная реконструкция таксономии ответственности позволяют сделать несколько наблюдений, важных для теории и практики ИИ-ориентированных госуслуг.

Во-первых, шестиуровневая пирамидальная модель автономности ИИ показывает (см. рис. 1), что ответственность не является бинарной, она распределена вдоль континуума делегирования решений алгоритмам и требует дифференцированного институционального подхода в зависимости от уровня автономии и критичности последствий (см. табл. 1). Анализ кейсов выявляет неоднородность ролей частных и публичных акторов и различия в правовых и политических последствиях ошибок систем.

Таблица 1

Таксономия классов ИИ и ее соотношение с уровнями рисков и критичностью компонента суверенного ИИ

Table 1. Taxonomy of AI classes and its relationship with risk levels and the criticality of the sovereign AI component

Класс ИИ	Функции	Участие человека	Уровень риска	Критичность компонента суверенного ИИ
ИИ-Капитан	Полное автоматическое принятие решений без участия человека	Отсутствует. Полная автономия ИИ при принятии решения. Человеческий фактор исключен полностью	Критически высокий (возможен ущерб жизням граждан, их благополучию и правам)	Максимальная угроза суверенитету: критическая зависимость от технологического партнера — поставщика ИИ при автономной выработке и принятии решений. Полностью зависит от поставщика: качество, полнота данных, релевантность данных для обучения ИИ, их полнота и непредвзятость, качество и транспарентность алгоритмов и кода. Зависимость от технической инфраструктуры
ИИ-Штурман	Генерация готового решения или сценария, окончательное решение принимают ответственные лица (утверждают/отклоняют)	Средний: человек утверждает или отклоняет предложение ИИ. Риск человеческой ошибки при принятии/отклонении сценария	Высокий (возможны ошибки в предлагаемых ИИ военных, стратегических и прочих сценариях)	Крайне высокая угроза суверенитету. Важен вопрос доверия к поставщику ИИ. Зависит от поставщика: качество, полнота данных, релевантность данных для обучения ИИ, их непредвзятость, прозрачность алгоритма и кода. Зависимость от технической инфраструктуры
ИИ-Советник	Предлагает варианты конфигураций и решений; выбор, корректировка и утверждение за ответственными лицами	Высокий: человек выбирает и дорабатывает решение. Риск человеческой ошибки при коррекции и утверждении сценариев	Ограниченный (непрозрачность / ошибки сценариев)	Высокая угроза суверенитету. Важно качество данных, на которых обучается ИИ, прозрачность алгоритмов. Инфраструктурная зависимость сохраняется

Окончание табл. 1

Класс ИИ	Функции	Участие человека	Уровень риска	Критичность компонента суверенного ИИ
ИИ-Наблюдатель	Автоматический сбор и анализ данных, обнаружение триггеров; контроль и реагирование остаются за ответственными лицами	Человек вырабатывает и принимает решение на основе сигналов ИИ. Возможны человеческие ошибки	Ограниченный (ущерб от ложных или запоздалых сигналов, ошибки данных)	Умеренное влияние на суверенитет. Зависимость от технологий мониторинга данных, алгоритмов обнаружения триггеров. Инфраструктурная зависимость сохраняется
ИИ-Рабочие руки	Выполнение сложных расчетных или трудозатратных задач без принятия ключевых решений	Человек контролирует и подтверждает результаты. Необходимость проверки. Возможность человеческой ошибки при контроле или настройке системы	Невысокий (формальные ошибки)	Менее критичен для суверенитета, однако массовое использование сторонних систем (например, провайдеров ML) создает долгосрочную технологическую зависимость. Также сохраняется инфраструктурная зависимость
ИИ-Рутинный помощник	Выполнение простых рутинных операций (перевод, сортировка, поддержка) под полным контролем ответственных лиц	Человек контролирует и управляет каждым этапом работы ИИ. Максимизация человеческого фактора	Минимальный	Невысокое влияние на суверенитет, но даже здесь использование сторонних ИИ способно привести к инфраструктурной зависимости

Источник: Составлено автором.

Во-вторых, анализ дилеммы государственно-частного партнерства показывает, что прагматизм госорганов, стремящихся быстро получить технологическое преимущество и сэкономить ресурсы, конфликтует с требованиями национальной безопасности, защиты данных и устойчивости инфраструктуры. Наличие готовых корпоративных решений облегчает внедрение, но создает системные уязвимости: зависимость от поставщика, риск утраты контроля над данными и технологиями, а также проблемы транспарентности, подотчетности и воспроизводимости алгоритмов. В этих условиях концепция «суверенного ИИ» выступает стратегической реакцией государств для восстановления контроля и автономии критических сервисов, хотя ее реализация сопряжена с серьезными ресурсными, организационными и нормативными препятствиями и не всегда достижима в краткосрочной перспективе. Важным наблюдением является факт того, что на всех уровнях сохраняется угроза суверенитету на технико-инфраструктурном уровне (хранение данных, аппаратные комплексы, вычислительные мощности и прочее) при реализации государственно-частного партнерства в области реализации ИИ-ориентированных сервисов.

Третье наблюдение касается институциональных пробелов в привлечении ответственности. Традиционные юридические и административные схемы, ориентированные на людей и организации, плохо адаптированы к гибридным ситуациям, где решения проходят через автоматизированный предсказательный модуль, рекомендательную подсистему и финальное утверждение человеком. Необходимы правовые и контрактные механизмы, распределяющие ответственность между разработчиком модели, провайдером инфраструктуры, оператором и контролирующим органом. Простое перекладывание ответственности на частного поставщика или пользователя не обеспечивает прозрачности и не снимает системные риски, что подтверждают рассмотренные кейсы.

Четвертый тезис раскрывает парадокс распределения ошибок в человеко-машинных системах принятия решений. На высоких уровнях автономии ИИ влияние человеческого фактора минимизируется, однако потенциальный ущерб от ошибки ИИ достигает максимума. И наоборот, снижение автономности системы повышает вероятность ошибки, обусловленной человеческим фактором (см. табл. 1). В данном контексте наиболее эффективными представляются системы среднего уровня в таксономии (см. рис. 1), в которых решения вырабатываются в процессе кооперации и симбиоза между человеком и ИИ, что позволяет взаимно компенсировать их слабые стороны и сбалансировать риски.

Анализ подтверждает необходимость многоуровневого управления рисками, включая пре- и поствнедренческую оценку критичности, обязательную верификацию и валидацию моделей для сценариев высокого и среднего риска, прозрачные процедуры аудита и институциональные гарантии человеческого контроля там, где ошибки влияют на права и безопасность граждан. Категории градации критичности демонстрируют теоретическую применимость, однако требуют адаптации к локальным институциональным и техническим реалиям на национальном и глобальном уровнях.

С теоретической точки зрения исследование подчеркивает важность перехода от чисто технологического дискурса к социотехнической аналитике, где ИИ рассматривается как часть сложной сети акторов, институтов и инфраструктур. Это требует объединения юридических, социальных и технических подходов в единый методологический каркас, способный быстро реагировать на изменения технологического поля. Методологические ограничения (качественный дизайн, выборка кейсов) влияют на воспроизводимость выводов, однако согласованность эмпирики и теоретической таксономии усиливает валидность предложенных интерпретаций в данных контекстах.

В итоге оценка показывает, что эффективное управление ИИ в госсекторе требует многоуровневой стратегии: от точной классификации сервисов по уровням автономии и рискам (см. табл. 1) до установления соответствующих правил ответственности и институтов контроля. Необходим дифференцированный подход к регуляции — строгие меры (обязательная сертификация, аудит) для систем высокого риска и высокой автономности и более гибкие требования для инструментов с низкой автономией. Такое разделение должно быть отражено и в институциональных механизмах, включая контракты с частными поставщиками, распределение юридической ответственности между разработчиком и ведомством, а также обеспечение прозрачной обратной связи с обществом. Международное сотрудничество в стандартизации и обмене опытом также повысит надежность решений для различных классов ИИ.

Выводы

Данная работа подчеркивает, что выработанная классификация ИИ-сервисов по уровням автономии (рис. 1, табл. 1) является ключевым элементом для грамотного управления их рисками и ответственностью. Конкретные рекомендации включают введение обязательной градации ИИ-систем по критичности задач и доле машинного участия, а также применение механизмов обеспечения прозрачности и объяснимости

для систем среднего и высокого риска. Необходимо внедрить процедуры обязательного постаудита и сертификации для наиболее автономных классов («Капитан», «Штурман»), а также разработать гибкие форматы международного сотрудничества по обмену практиками регулирования и стандартов. Эти элементы не исчерпывают все меры, но формируют прагматическую дорожную карту для снижения рисков и укрепления подотчетности при масштабном внедрении ИИ в публичный сектор.

Ключевой вывод исследования заключается в том, что внедрение ИИ-ориентированных государственных сервисов — это не только техническая модернизация, но и глубокая институциональная и политическая трансформация публичного управления. Эффективное управление требует сочетания юридических реформ, публичного, экспертного и политического дискурса, институциональной адаптации и международной кооперации для создания устойчивых, контролируемых и подотчетных ИИ-систем. Без системного подхода к таксономии ответственности и ролей ИИ, балансу интересов в государственно-частном партнерстве и разработке национальных стратегий суверенитизации ИИ риск технологической зависимости и институциональной растерянности возрастает, угрожая правам граждан и результативности государственных сервисов.

Литература/References

1. Al-Suqri M., Niaz H. A comparative analysis of information and artificial intelligence toward national security // Ieee Access. 2022. Vol. 10. P. 64420–64434. DOI 10.1109/access.2022.3183642.
2. Barth T. J., Arnold E. Artificial intelligence and administrative discretion: Implications for public administration // The American Review of Public Administration. 1999. Vol. 29. N 4. P. 332–351. DOI 10.1177/02750749922064463.
3. Braun V., Clarke V. Using thematic analysis in psychology // Qualitative Research in Psychology. 2006. Vol. 3, N 2. P. 77–101.
4. Chia H., et al. Autonomous AI: what does autonomy mean in relation to persons or machines? // Law, Innovation and Technology. 2023. Vol. 15. N 2. P. 390–410. DOI 10.1080/17579961.2023.2245679.
5. Criado J. I., Sandoval-Almazбn R., Gil-Garcia J. R. Artificial intelligence and public administration: Understanding actors, governance, and policy from micro, meso, and macro perspectives // Public Policy and Administration. 2025. Vol. 40. N 2. P. 173–184. DOI 10.1177/09520767241272921.
6. Dale R. Sovereign AI in 2025 // Natural Language Processing. 2025. P. 1–10.
7. Datta K. AI-driven public administration: Opportunities, challenges, and ethical considerations // The Social Science Review. 2024. Vol. 2. N 6. P. 134–139. DOI 10.70096/tssr.240206023.
8. de Souza E. A. The Transformation of Public Administration through Artificial Intelligence // Rev. fisio&terapia. 2025. Vol. 29. N 145. P. 44–45.
9. Dillon S., Dillon M. Artificial intelligence and the sovereign-governance game. In book: AI Narratives. 2020. P. 333–356. DOI 10.1093/oso/9780198846666.003.0015.
10. Djeffal C., Siewert M. B., Wurster S. Role of the state and responsibility in governing artificial intelligence: a comparative analysis of AI strategies // Journal of European Public Policy. 2022. Vol. 29. N 11. P. 1799–1821. DOI 10.1080/13501763.2022.2094987.
11. Filgueiras F. Designing artificial intelligence policy: comparing design spaces in Latin America // Latin American Policy. 2023. Vol. 14. N 1. P. 5–21. DOI 10.1111/lamp.12282.
12. Henman P. Improving public services using artificial intelligence: possibilities, pitfalls, governance // Asia Pacific Journal of Public Administration. 2020. Vol. 42. N 4. P. 209–221. DOI 10.1080/23276665.2020.1816188.
13. Hou Y., Wang Z., Yang Z., Zhai E. Artificial intelligence technology pushes forward the modernization of firepower weapon equipmen // Proc. SPIE 12720, 2022 Workshop on Electronics Communication Engineering, 127200F (28 June 2023). https://doi.org/10.1117/12.2668167.
14. Khanal S., Zhang H., Taeihagh A. Why and how is the power of Big Tech increasing in the policy process? The case of generative AI // Policy and Society. 2025. Vol. 44, N 1. P. 52–69. DOI 10.1093/polsoc/puae012.
15. Kurki V. The legal personhood of artificial intelligences. In book: A Theory of Legal Personhood. 2019. P. 175–190. DOI 10.1093/oso/9780198844037.003.0007.
16. Li C. AI-Driven Governance: Enhancing Transparency and Accountability in Public Administration // Digital Society & Virtual Governance. 2025. Vol. 1. N 1. P. 1–16. DOI 10.6914/dsvg.010101.

17. Liu L. X., Clegg S., Pollack J. The Effect of Public–Private Partnerships on Innovation in Infrastructure Delivery // Project Management Journal. 2024. Vol. 55. N 1. P. 31–49. DOI 10.1177/87569728231189989.
18. Madan R. Artificial intelligence diffusion in public administration // Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. 2022. P. 903. DOI 10.1145/3514094.3539529.
19. Ortega-Fernбndez A., Martнn-Rojas R., Garcна-Morales V. J. Artificial Intelligence in the Urban Environment: Smart Cities as Models for Developing Innovation and Sustainability // Sustainability. 2020. Vol. 12. N 19. P. 7860. DOI 10.3390/SU12197860.
20. Prasad K. R., Karanam S. R., Ganesh D., Liyakat K. K. S., Talasila V., Purushotham P. AI in public-private partnership for IT infrastructure development // The Journal of High Technology Management Research. 2024. Vol. 35. N 1. P. 100496. DOI 10.1016/j.hitech.2024.100496.
21. Pulijala S. Artificial intelligence in governance: opportunities, challenges, and ethical implications for public administration // International Journal for Multidisciplinary Research (IJFMR). 2024. Vol. 6. N 6. P. 1–10. DOI 10.36948/ijfmr.2024.v06i06.29990.
22. Roberts H. Digital sovereignty and artificial intelligence: a normative approach // Ethics Inf Technol. 2024. Vol. 26. P. 70. DOI 10.1007/s10676-024-09810-5.
23. Sen A. Artificial intelligence and autonomous systems: A legal perspective on granting personhood and implications of such a decision // DME Journal of Law. 2023. Vol. 4. N 01. P. 15–26. DOI 10.53361/dmejl.v4i01.03.
24. Taddeo M., McNeish D., Blanchard A., Edgar E. Ethical principles for artificial intelligence in national defence // Philosophy & Technology. 2021. Vol. 34, N 4. P. 1707–1729. DOI 10.1007/s13347-021-00482-3.
25. Timmers P. Ethics of AI and Cybersecurity When Sovereignty is at Stake // Minds & Machines. 2019. Vol. 29. P. 635–645. DOI 10.1007/s11023-019-09508-4.
26. Trajkovski G. Bridging the public administration-AI divide: A skills perspective // Public Administration and Developmen. 2024. Vol. 44. N 5. P. 412–426. DOI 10.1002/pad.2061.
27. van Noordt C., Tangi L. The dynamics of AI capability and its influence on public value creation of AI within public administration // Government Information Quarterly. 2023. Vol. 40. N 4. P. 101860. DOI 10.1016/j.giq.2023.101860.
28. Vatamanu A. F., Tofan M. Integrating artificial intelligence into public administration: Challenges and vulnerabilities // Administrative Sciences. 2025. Vol. 15. N 4. P. 149.
29. Webster J., Watson R. Analyzing the Past to Prepare for the Future: Writing a Literature Review // MIS Quarterly. 2002. Vol. 26. N 2. P. xiii–xxiii.
30. Zhang H., Khanal S., Taeihagh A. Public-Private Powerplays in Generative AI Era: Balancing Big Tech Regulation Amidst Global AI Race // Digital Government: Research and Practice. 2025. Vol. 6. N 2. P. 1–11. DOI 10.1145/3664824.

Конфликт интересов
Автор заявляет об отсутствии конфликта интересов.

Об авторе:
Носиков Андрей Андреевич, кандидат политических наук, старший преподаватель кафедры связей с общественностью в политике и государственном управлении Института «Высшая школа журналистики и массовых коммуникаций» Санкт-Петербургского государственного университета (Санкт-Петербург, Российская Федерация); a.nosikov@spbu.ru

Conflict of interests
The author declares no relevant conflict of interests.

About the author:
Andrey A. Nosikov, Ph.D. in Political Sciences, Senior Lecturer of Department of Public Relations in Politics and Public Administration, Institute «Higher School of Journalism and Mass Communications» of Saint Petersburg State University (Saint Petersburg, Russian Federation); a.nosikov@spbu.ru

Поступила в редакцию: 28.08.2025
Поступила после рецензирования: 04.10.2025
Принята к публикации: 15.10.2025

The article was submitted: 28.08.2025
Approved after reviewing: 04.10.2025
Accepted for publication: 15.10.2025