

**Информация для цитирования:**

Мартынов А. В., Ширеева Е. В. Правовые и практические вопросы использования синтетических данных для целей государственного управления в России // Вестник Пермского университета. Юридические науки. 2025. Вып. 4(70). С. 526–538. DOI: 10.17072/1995-4190-2025-70-526-538.

Martynov A. V., Shireeva E. V. *Pravovye i prakticheskie voprosy ispol'zovaniya sinteticheskikh dannykh dlya tseley gosudarstvennogo upravleniya v Rossii* [Legal and Practical Issues of Synthetic Data Use for Public Administration Purposes in Russia]. *Vestnik Permskogo universiteta. Juridicheskie nauki* – Perm University Herald. Juridical Sciences. 2025. Issue 4(70). Pp. 526–538. (In Russ.). DOI: 10.17072/1995-4190-2025-70-526-538.

УДК 342

DOI: 10.17072/1995-4190-2025-70-526-538

## Правовые и практические вопросы использования синтетических данных для целей государственного управления в России

Исследование выполнено за счет гранта Российского научного фонда № 25-28-00491,  
<https://rscf.ru/project/25-28-00491/>

**А. В. Мартынов**

Национальный исследовательский  
Нижегородский государственный университет  
имени Н. И. Лобачевского  
E-mail: docpred@yandex.ru

**Е. В. Ширеева**

Национальный исследовательский  
Нижегородский государственный университет  
имени Н. И. Лобачевского  
E-mail: shireevaekaterina@yandex.ru

Статья поступила в редакцию 10.07.2025

**Введение:** данные становятся основным стратегическим ресурсом, без которого невозможен качественный переход к развитому генеративному искусственному интеллекту и совершенствование методов и подходов к работе с большими данными. Однако обеспечение доступа к большим объемам данных широкого круга пользователей сталкивается с рядом существенных и очень устойчивых правовых барьеров. **Цель** настоящего исследования – определить практическое значение публичных цифровых озер для создания синтетических данных, а также рассмотреть перспективы их использования для целей публичного управления. **Методы:** исследование базируется на совокупности общенаучных (анализ, синтез, индукция, дедукция), частнонаучных (статистический, социологический) и специальных методов познания (формально-юридический, сравнительно-правовой, метод правового моделирования). **Результаты и выводы:** ключевым вопросом становится определение прав в отношении данных (установление обладателя данных) и наличия в массивах данных чувствительной информации, общий доступ к которой, в силу ряда причин, невозможен и несет в себе существенные риски для государства, общества или личности. В рамках этой тенденции оптимальным решением видится формирование крупных массивов данных (озер данных), управляемых государством, а также распространение практики использования синтетических данных. Россия на пути к экономике данных активно формирует публичные озера данных, которые в потенциале могут стать основой для создания на их основе синтетических данных. С этой целью был проанализирован российский и зарубежный опыт правового регулирования в указанной сфере, выявлены и сформулированы практические и правовые проблемы применения цифровых озер для синтетических данных в целях публичного управления.

© Мартынов А. В., Ширеева Е. В., 2025



Данная работа распространяется по лицензии CC BY 4.0. Чтобы просмотреть копию этой лицензии, посетите <https://creativecommons.org/licenses/by/4.0/>

**Information for citation:**

Martynov A. V., Shireeva E. V. *Pravovye i prakticheskie voprosy ispol'zovaniya sinteticheskikh dannykh dlya tseley gosudarstvennogo upravleniya v Rossii* [Legal and Practical Issues of Synthetic Data Use for Public Administration Purposes in Russia]. *Vestnik Permskogo universiteta. Juridicheskie nauki* – Perm University Herald. Juridical Sciences. 2025. Issue 4(70). Pp. 526–538. (In Russ.). DOI: 10.17072/1995-4190-2025-70-526-538.

Мартынов А. В., Ширеева Е. В. Правовые и практические вопросы использования синтетических данных для целей государственного управления в России // Вестник Пермского университета. Юридические науки. 2025. Вып. 4(70). С. 526–538. DOI: 10.17072/1995-4190-2025-70-526-538.

UDC 342

DOI: 10.17072/1995-4190-2025-70-526-538

## Legal and Practical Issues of Synthetic Data Use for Public Administration Purposes in Russia

The study was supported by a grant from the Russian Science Foundation, project No. 25-28-00491,  
<https://rscf.ru/project/25-28-00491/>

**A. V. Martynov**

Lobachevsky State University of Nizhny Novgorod  
E-mail: docpred@yandex.ru

**E. V. Shireeva**

Lobachevsky State University of Nizhny Novgorod  
E-mail: shireevaekaterina@yandex.ru

Received 10 Jul 2025

**Introduction:** data are becoming the main strategic source, without which it is impossible to make a qualitative transition to advanced generative artificial intelligence and improve methods and approaches to big data processing. However, there are some significant and deep-rooted legal barriers on the way to providing access to large amounts of data for a wide range of users. **Purpose:** the study aims to substantiate the practical importance of public data lakes for creating synthetic data as well as to explore the prospects for their use for public administration purposes. **Methods:** the research is based on a combination of general scientific methods (analysis, synthesis, induction, deduction), special scientific methods (statistical, sociological), and specialized methods of cognition (formal legal, comparative legal methods, legal modeling). **Results and conclusions:** the key issues are the rights in relation to data (who owns the data) and the presence in datasets of sensitive information that, for a number of reasons, must not be openly shared and carries significant risks for the state, society, or the individual. What seems to be the most optimal solution is the formation of large arrays of data (data lakes) managed by the state as well as the spread of the practice of using synthetic data. On its way to a data economy, Russia is actively developing the practice of creation of public data lakes, which can potentially become the source for creating synthetic data. The paper analyzes Russian and foreign experience of legal regulation in this area, identifying practical and legal issues of the use of synthetic data for public administration purposes.

© Martynov A. V., Shireeva E. V., 2025



This work is licensed under CC BY 4.0. To view a copy of the license, visit <https://creativecommons.org/licenses/by/4.0/>

**Ключевые слова:** синтетические данные; озеро данных; государственное управление; органы исполнительной власти; искусственный интеллект; правовое регулирование; генеративный искусственный интеллект; машинное обучение

## Legal and Practical Issues of Synthetic Data Use for Public Administration Purposes in Russia

The study was supported by a grant from the Russian Science Foundation, project No. 25-28-00491, <https://rscf.ru/project/25-28-00491/>

**A. V. Martynov**

Lobachevsky State University of Nizhny Novgorod  
E-mail: docpred@yandex.ru

**E. V. Shireeva**

Lobachevsky State University of Nizhny Novgorod  
E-mail: shireevaekaterina@yandex.ru

Received 10 Jul 2025

**Introduction:** data are becoming the main strategic source, without which it is impossible to make a qualitative transition to advanced generative artificial intelligence and improve methods and approaches to big data processing. However, there are some significant and deep-rooted legal barriers on the way to providing access to large amounts of data for a wide range of users. **Purpose:** the study aims to substantiate the practical importance of public data lakes for creating synthetic data as well as to explore the prospects for their use for public administration purposes. **Methods:** the research is based on a combination of general scientific methods (analysis, synthesis, induction, deduction), special scientific methods (statistical, sociological), and specialized methods of cognition (formal legal, comparative legal methods, legal modeling). **Results and conclusions:** the key issues are the rights in relation to data (who owns the data) and the presence in datasets of sensitive information that, for a number of reasons, must not be openly shared and carries significant risks for the state, society, or the individual. What seems to be the most optimal solution is the formation of large arrays of data (data lakes) managed by the state as well as the spread of the practice of using synthetic data. On its way to a data economy, Russia is actively developing the practice of creation of public data lakes, which can potentially become the source for creating synthetic data. The paper analyzes Russian and foreign experience of legal regulation in this area, identifying practical and legal issues of the use of synthetic data for public administration purposes.

**Keywords:** synthetic data; data lake; public administration; executive authorities; artificial intelligence; legal regulation; generative artificial intelligence; machine learning

### Введение

Развитие технологии искусственного интеллекта (ИИ) высокого уровня порождает необходимость получения и обработки огромного массива информации. Она используется для обучения искусственного интеллекта как в ручном, так и в автономном режиме. При этом образуемые нейросети функционируют не только на основе существующей информации, полученной разными способами и через различные источники, но и посредством создания новых данных, которые не имеют естественного (исходящего от человека) происхождения.

На современном этапе развития цифровых технологий именно генеративный искусственный интеллект считается искусственным интеллектом высокого уровня, то есть способным преобразовывать существующую информацию и создавать абсолютно новые данные. Эти данные (информация) наиболее приближены к реальным данным (информации),

создаваемым человеком (литературное произведение, служебный документ, картина, фотография, изображение и т. п.). При этом обеспечивается необходимый уровень достоверности новых данных, что свидетельствует об их приемлемости и допустимости в различных общественных процессах.

Речь идет об информации, созданной или преобразованной генеративным искусственным интеллектом. Такая информация (данные, сведения, цифровые объекты) получила наименование синтетических данных. В самом упрощенном виде под синтетическими данными обычно понимают информацию, сгенерированную компьютерами и по своим статистическим свойствам напоминающую реальные данные. Она создается с помощью алгоритмов без прямой связи с реальными сведениями [15].

Илон Маск в интервью газете «Гардиан» заявил, что у компаний, занимающихся искусственным интеллектом, закончились данные для обучения их моделей и что они «исчерпали» все человеческие

**Keywords:** synthetic data; data lake; public administration; executive authorities; artificial intelligence; legal regulation; generative artificial intelligence; machine learning

### Information in Russian

## Правовые и практические вопросы использования синтетических данных для целей государственного управления в России

Исследование выполнено за счет гранта Российского научного фонда № 25-28-00491, <https://rscf.ru/project/25-28-00491/>

**А. В. Мартынов**

Национальный исследовательский  
Нижегородский государственный университет  
имени Н. И. Лобачевского  
E-mail: docpred@yandex.ru

**Е. В. Ширеева**

Национальный исследовательский  
Нижегородский государственный университет  
имени Н. И. Лобачевского  
E-mail: shireevaekaterina@yandex.ru

Статья поступила в редакцию 10.07.2025

**Введение:** данные становятся основным стратегическим ресурсом, без которого невозможен качественный переход к развитому генеративному искусственному интеллекту и совершенствование методов и подходов к работе с большими данными. Однако обеспечение доступа к большим объемам данных широкого круга пользователей сталкивается с рядом существенных и очень устойчивых правовых барьеров. **Цель** настоящего исследования – определить практическое значение публичных цифровых озер для создания синтетических данных, а также рассмотреть перспективы их использования для целей публичного управления. **Методы:** исследование базируется на совокупности общенаучных (анализ, синтез, индукция, дедукция), частнонаучных (статистический, социологический) и специальных методов познания (формально-юридический, сравнительно-правовой, метод правового моделирования). **Результаты и выводы:** ключевым вопросом становится определение прав в отношении данных (установление обладателя данных) и наличия в массивах данных чувствительной информации, общий доступ к которой, в силу ряда причин, невозможен и несет в себе существенные риски для государства, общества или личности. В рамках этой тенденции оптимальным решением видится формирование крупных массивов данных (озер данных), управляемых государством, а также распространение практики использования синтетических данных. Россия на пути к экономике данных активно формирует публичные озера данных, которые в потенциале могут стать основой для создания на их основе синтетических данных. С этой целью был проанализирован российский и зарубежный опыт правового регулирования в указанной сфере, выявлены и сформулированы практические и правовые проблемы применения цифровых озер для синтетических данных в целях публичного управления.

**Ключевые слова:** синтетические данные; озеро данных; государственное управление; органы исполнительной власти; искусственный интеллект; правовое регулирование; генеративный искусственный интеллект; машинное обучение

### Introduction

The development of high-level artificial intelligence (AI) technology necessitates the acquisition and processing of vast amounts of data. This information is utilized to train AI systems, both manually and through offline methods. The neural networks formed operate not only on the basis of existing (real) information, obtained through various means and sources, but also through generation of new data, this lacking a natural (i.e., human-generated) origin.

At the current stage of the digital technology development, generative AI is regarded as a form of advanced artificial intelligence. This technology is capable of transforming existing information and creating entirely new data. Generated data closely resembles

authentic human-created content, such as literary works, official documents, paintings, photographs, and images. Furthermore, the information produced maintains a necessary level of reliability, indicating its trustworthiness and suitability for various social processes.

Here, we are speaking about information created or transformed by generative AI. This type of information (data, digital objects) has received the name 'synthetic data'. In its simplest form, synthetic data is understood as information generated by computers that statistically resembles real data. It is created by algorithms, without a direct connection to actual records [15].

In an interview with *The Guardian*, Elon Musk stated that IT companies have exhausted their data resources for training AI models, claiming they have of all

знания. Он предположил, что технологическим компаниям придется обратиться к синтетическим данным – или материалам, созданным моделями искусственного интеллекта, – чтобы создавать и настраивать новые системы [14]. Эндрю Дункан, директор по науке и инновациям (фундаментальные исследования в области DS и AI) в британском Институте Алана Тьюринга, подтвердил, что комментарий Маска согласуется с недавней научной работой, в которой говорится, что общедоступные данные для моделей ИИ могут закончиться уже в 2026 году [17]. Он добавил, что чрезмерная зависимость от синтетических данных способна привести к «краху модели» – так называют ухудшение качества результатов работы моделей [14]. Вместе с тем уже сейчас могут быть применены весьма эффективные решения для преодоления этой проблемы. Речь идет о цифровых озерах данных (Data Lake) как источнике данных для обучения генеративного искусственного интеллекта.

Озеро данных – это, как правило, единое хранилище данных, включающее необработанные копии данных исходной системы, данные датчиков, социальные данные и т. д., а также преобразованные данные, используемые для составления отчетов, визуализации, расширенной аналитики и машинного обучения и т. п.

#### Методологическая основа исследования

Исследование базируется на совокупности общенаучных (анализ, синтез, индукция, дедукция), частнонаучных (статистический, социологический) и специальных методов познания (формально-юридический, сравнительно-правовой, метод правового моделирования). Поскольку тема исследования имеет правовую специфику, остановимся на характеристике лишь специальных методов познания, которые присущи исключительно юриспруденции. Формально-юридический метод применялся для уточнения содержания норм права: он позволил их истолковать, определить структуру и терминологию законов. Кроме того, с его помощью были выявлены теоретические и практические проблемы применения синтетических данных и цифровых озер в деятельности публичных органов. Сравнительно-правовой метод использовался для соотнесения правового регулирования и практики применения синтетических данных и цифровых озер в публичном секторе в России, США, Европе и Китае. Он позволил выделить критерии для сравнения правового регулирования, выявить специфические и общие черты в практике применения публичными органами цифровых озер и синтетических данных. Поскольку в исследовании определены проблемы научного и правового характера, применение метода правового моделирования позволило сформулировать решение этих проблем, а также выводы по научному исследованию в целом.

#### Теоретическая основа исследования

Следует отметить, что исследования о правовом регулировании применения цифровых озер и синтетических данных в публичном секторе фактически отсутствуют. Это связано с тем, что данные технологии возникли в публичном секторе большинства государств лишь в 2022–2024 годах. Есть лишь исследования по некоторым аспектам применения цифровых озер и синтетических данных.

Теоретические основы цифровых озер описаны Джеймсом Диксоном как альтернатива «хранилищу данных» [8]. Структура цифровых озер и их технические принципы работы в публичном секторе рассмотрены в исследованиях Билла Инмона [10]. Роль цифровых озер в деятельности органов государственной власти изучена Робом Китчином [12], а также Инес Мергель, Ноэллой Эдельманн и Натали Хауг [13]. Имеются, кроме того, исследования, посвященные опыту применения цифровых озер в конкретных государствах [1; 6; 9; 18]. Вопросы неприкосновенности частной жизни и риски нарушения конфиденциальности тоже становились предметом исследования [7; 9; 18].

Настоящее исследование освещает не только вышеназванные аспекты. Его спецификой является рассмотрение вопроса через призму проблематики публичного управления с учетом правового опыта и практики применения цифровых озер и синтетических данных в Российской Федерации.

#### Международный опыт. Регулирование синтетических данных и озер данных. Искусственный интеллект

##### 1. США

Использование озер данных государственными структурами в США имеет довольно давнюю историю. Реализованные инициативы и решения были направлены на создание или улучшение использования правительственными агентствами больших массивов данных.

С учетом этого появляется все больше проектов, направленных на интеграцию разрозненных хранилищ данных в единые правительственные озера данных, а также привлечение к этому частных компаний для выполнения подрядных работ. Примером могут служить инициативы Министерства обороны США (US Department of Defense) и Бюро переписи населения США (US Census Bureau), которые в 2024 году объявили о серьезных планах интеграции структурированных, полуструктурированных и «сырых» озер данных [6]. Возрастает у федеральных агентств США интерес и к созданию публичных, то есть открытых, озер данных, которые являются своего рода продолжением уже давно осознанной тенденции обеспечения открытости данных. В частности, это подтверждает создание Public-Use Data Lake, инициированное Министерством труда США (US Department of Labour).

human knowledge. He suggested that these companies would need to turn to synthetic data – materials generated by AI models – to develop and configure new systems [14]. Andrew Duncan, Director of Science & Innovation (Fundamental Research in DS and AI) at the Alan Turing Institute in Britain, noted that Musk's comments align with recent research indicating that publicly available data for AI models could be depleted as early as 2026 [17]. He cautioned that an overreliance on synthetic data could lead to the deterioration in the quality of results produced by these models [14]. Meanwhile, effective solutions are already being implemented to address this issue: digital data lakes are being utilized as a source for training generative AI.

A data lake is typically a centralized data storage repository that contains raw copies of data from various source systems, including sensor data, social media data, and more, and also includes transformed data utilized for tasks such as reporting, visualization, advanced analytics, machine learning, etc.

#### Research Methodology

The research is grounded in a combination of general scientific methods (analysis, synthesis, induction, deduction), private scientific methods (statistical, sociological), and specialized cognitive methods (formal legal, comparative legal, and legal modeling). Given the legal specificity of the research topic, we will concentrate on the characteristics of the specialized cognitive methods that are unique to jurisprudence. The formal legal method was employed to clarify the content of legal norms, enabling the interpretation, structuring, and terminology of laws. Additionally, it was utilized to identify both theoretical and practical challenges associated with the use of synthetic data and data lakes in the operations of public authorities. The comparative legal method was employed to analyze the legal regulations and practices surrounding the use of synthetic data and data lakes in the public sectors of Russia, the USA, Europe, and China. This approach enabled us to establish criteria for comparing legal frameworks and to identify both unique and shared characteristics in the utilization of data lakes and synthetic data by public authorities. Since the study addresses issues of both scientific and legal significance, the application of legal modeling facilitated the development of solutions to these challenges as well as the formulation of broader conclusions regarding the scientific inquiry as a whole.

#### Literature Review

It is important to note that there are currently almost no studies on the legal regulation of digital lakes and synthetic data within the public sector. This lack of research can be attributed to the fact that these technologies were only introduced in the public sectors of most countries between 2022 and 2024. Furthermore, existing studies focus on only certain aspects of the use of digital lakes and synthetic data.

The theoretical foundations of digital lakes are described by James Dixon as an alternative to the traditional 'data warehouse' [8]. The structure of digital lakes and their technical principles of operation within the public sector are examined in the research conducted by W. H. Inmon [10]. Additionally, the role of digital lakes in the activities of public authorities is analyzed by R. Kitchin [12] and I. Mergel, N. Edelman, and N. Haug [13].

There are also studies focused on examining the experience of using digital lakes in specific states [1], [6], [9], [18]. Additionally, privacy and confidentiality risks have been partially addressed in research [7], [9], [18].

The present study addresses not only the aforementioned aspects. It is special in that it examines the issue through the lens of public administration and with the consideration of the legal experience and practice of application of digital lakes and synthetic data in the Russian Federation.

#### International Practice. Regulation of Synthetic Data and Data Lakes. Artificial Intelligence

##### 1. USA

The use of data lakes by government agencies in the United States has a long history. Numerous initiatives and decisions have been implemented to create or enhance the utilization of large volumes of data by government entities.

In this context, an increasing number of projects are focused on integrating disparate data warehouses into unified government data lakes, while also engaging private companies for contract work. A notable example is the initiative by the U.S. Department of Defense and the U.S. Census Bureau that announced in 2024 comprehensive plans to integrate structured, semi-structured, and raw data lakes [6]. There is a growing interest among U.S. federal agencies in creating publicly accessible data lakes, which are, to some extent, a continuation of the longstanding trend of promoting data transparency. An example of this is the Public-Use Data Lake established by the U.S. Department of Labor.

Современные потребности в применении технологий генеративного искусственного интеллекта в государственном управлении как на федеральном уровне, так и на уровне штатов диктуют необходимость крайне разборчиво подходить к выбору этих технологий по причине высоких рисков их влияния одновременно на интересы государственного управления и обеспечение прав граждан. Сегодня уже очевидно, что применение реальных данных грозит раскрытием чувствительной информации и незаконным ее использованием с учетом существующих ограничений в доступе [18]. Наиболее вероятным выходом из ситуации представляется использование для тестирования рассматриваемых к применению технологий генеративного ИИ искусственных данных и озер данных.

В частности, уже есть конкретные примеры сотрудничества местных правительств с университетскими исследовательскими центрами и частными компаниями. Так, Центр системных данных Мэриленда (Maryland Longitudinal Data System Center) сотрудничает с несколькими исследовательскими центрами для формирования на основе образовательных данных по школьникам и студентам округа искусственных данных для последующего их использования в целях улучшения публичных услуг в этой сфере. Другим примером может служить сотрудничество Института городского развития (Urban Institute) с Департаментом социальных служб округа Аллегейни (Allegheny County Department of Human Services) и Региональным центром данных Западной Пенсильвании (Western Pennsylvania Regional Data Center) в рамках пилотного проекта по использованию синтетических данных для улучшения социальных слуг и программ [16].

На федеральном уровне также отмечен интерес к проблематике использования искусственных данных в различных целях. В январе 2024 года Федеральный совет руководителей служб обработки данных (Federal Chief Data Officers Council) опубликовал запрос на сотрудничество в проведении исследований в области искусственных данных, отметив, что современное понимание этого явления и подходы к нему не отвечают актуальным потребностям государства и бизнеса в развитии и регулировании искусственных данных, имеющих необходимый потенциал для экономики.

## 2. Европа

В Европейском союзе присутствует выраженный интерес к правовому регулированию вопросов использования технологий искусственного интеллекта и оборота данных в сфере публичного управления. В этой связи в дополнение к уже широко известному Общему регламенту по защите данных (General Data Protection Regulation)<sup>1</sup> были приняты Закон об искусственном интеллекте (Artificial Intelligence Act)<sup>2</sup> и Закон о данных (Data Act)<sup>3</sup>, которые в совокупности определяют общую стратегию для стран Европы как в части управления данными, так и в части использования технологий ИИ.

Так, Закон об искусственном интеллекте содержит упоминание о том, что не существует непреодолимых препятствий для развития, внедрения и использования технологий искусственного интеллекта в сфере государственного управления и предоставления публичных услуг. Однако при этом должны быть соблюдены все необходимые фундаментальные права и свободы личности. В частности, Европейский надзорный орган по защите данных (European Data Protection Supervisor) в 2023 году опубликовал первый набор рекомендаций (Guidelines) по использованию технологий генеративного искусственного интеллекта для органов и учреждений Евросоюза<sup>4</sup>. Несмотря на то что документ в принципе имеет рекомендательный характер, он содержит вполне четкие и конкретные правила, которые позволяют провести оценку потенциальных рисков применения указанных технологий.

Персонал органов Евросоюза должен придерживаться как минимум пяти важных принципов при использовании генеративного искусственного интеллекта:

- служащий не вправе разглашать и передавать для использования генеративным искусственным интеллектом информацию, которая не имеет статуса общедоступной или публичной;
- следует критично относиться к полученному результату в части его точности и релевантности, а также наличия в нем противоречий или ложной информации;
- следует критически подходить к использованию публичных генеративных технологий искусственного

<sup>1</sup> General Data Protection Regulation. URL: <https://eur-lex.europa.eu/EN/legal-content/summary/general-data-protection-regulation-gdpr.html>.

<sup>2</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>.

<sup>3</sup> Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). URL: <https://eur-lex.europa.eu/eli/reg/2023/2854>.

<sup>4</sup> First EDPS Orientations for ensuring data protection compliance when using Generative AI systems, European Data Protection Supervisor. URL: [https://www.edps.europa.eu/system/files/2024-06/24-06-03\\_genai\\_orientations\\_en.pdf](https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf).

## 2. Europe

In the European Union, there is a growing interest in the legal regulation of artificial intelligence (AI) technologies and data processing for the benefit of public administration. In this context, in addition to the well-known General Data Protection Regulation (GDPR)<sup>1</sup>, the EU has adopted the Artificial Intelligence Act<sup>2</sup> and the Data Act<sup>3</sup>. Together, these regulations define a cohesive strategy for European countries with regard to both data management and AI technology utilization.

The EU Artificial Intelligence Act specifically states that there are no significant barriers to the development, implementation, and use of AI technologies in public administration and the provision of public services. However, it is essential to respect all fundamental rights and freedoms of individuals. In 2023, the European Data Protection Supervisor published the first set of guidelines on the use of generative AI technologies for EU authorities and institutions<sup>4</sup>. Although the document is advisory in nature, it provides clear and specific rules for assessing potential risks associated with the use of generative AI technologies.

The staff of the EU's authorities should adhere to at least five key principles when utilizing generative AI:

- an employee does not have the right to disclose or transmit to generative AI information that is not publicly available or has no open access;
- the results obtained from generative AI should be critically evaluated for accuracy and relevance, as well as for the presence of contradictions or false information;
- it is essential to adopt a critical approach to the use of publicly available (open-source) generative AI

The contemporary demands for generative AI technologies in public administration, at both federal and state levels, necessitate a careful selection of these technologies due to the significant risks they pose to public interests and the protection of citizens' rights. In this context, the use of real data raises concerns about the potential disclosure of sensitive information and its misuse, in circumvention of the existing restrictions on data access [18]. A promising solution can be seen in the use of synthetic data and data lakes for testing generative AI technologies.

Today it is possible to give several examples of collaboration among local governments, university research centers, and private companies. The Maryland Longitudinal Data System Center is currently working with various research centers to generate synthetic data derived from educational information about schoolchildren and students in the district. The synthetic data obtained will be utilized for the purpose of enhancing public services in this field. Another example is the Urban Institute's partnership with the Allegheny County Department of Human Services and the Western Pennsylvania Regional Data Center, which aims to pilot research on the use of synthetic data to improve social services and programs [16].

At the federal level, there is a pressing need to examine the challenges associated with the use of synthetic data for various applications. In January 2024, the Federal Chief Data Officers Council published a request for collaboration to conduct research in the field of synthetic data, noting that the current definitions and approaches did not adequately address the actual needs of both the government and businesses in terms of the development and regulation of synthetic data as data possessing significant potential for the economy.

<sup>1</sup> General Data Protection Regulation. Available at: <https://eur-lex.europa.eu/EN/legal-content/summary/general-data-protection-regulation-gdpr.html>

<sup>2</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>

<sup>3</sup> Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). Available at: <https://eur-lex.europa.eu/eli/reg/2023/2854>

<sup>4</sup> First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. *European Data Protection Supervisor* [website]. Available at: [https://www.edps.europa.eu/system/files/2024-06/24-06-03\\_genai\\_orientations\\_en.pdf](https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf).

интеллекта во избежание нарушения прав на интеллектуальную собственность;

– нельзя использовать генеративный искусственный интеллект для создания юридически значимых или обязывающих документов;

– нельзя использовать генеративный искусственный интеллект для критических и «чувствительных» процессов.

Сам по себе Закон об искусственном интеллекте содержит ряд важных правил использования искусственного интеллекта для целей публичного управления, а именно связанных с применением особой маркировки как для результатов, полученных с помощью искусственного интеллекта, так и для случаев взаимодействия гражданина и искусственного интеллекта при оказании публичных услуг.

Закон о данных также внес серьезные коррективы в регулирование оборота данных в рамках Европейского союза. С учетом целей и задач настоящего исследования необходимо отметить то, что данный документ предоставил органам власти право получить доступ к данным, используемым и накапливаемым субъектами частного сектора, при определенных условиях чрезвычайных ситуаций.

Еще одним документом, который активно влияет на создание рынка данных в Европейском союзе, является Директива об открытых данных (Open Data Directive)<sup>1</sup>, которая четко закрепляет необходимость раскрытия информации, собираемой органами власти по шести тематическим областям (геопространственная; наблюдение за Землей и окружающей средой; метеорологическая; статистическая; компании и собственность компаний; мобильность). Наборы данных должны предоставляться бесплатно в машиночитаемых форматах и через API для облегчения их эффективного повторного использования. В целом реализация Директивы привела к появлению и развитию European Data Portal – общеевропейской многоязычной платформы для поиска информации государственного сектора, открытой для повторного использования.

В рамках рассматриваемой стратегии, в части построения рынка данных в Европейском союзе, синтетические данные все чаще рассматриваются как аналог анонимизированных данных для снижения риска нанесения ущерба правам и интересам физических лиц при обработке персональных данных [9], о чем более подробно изложено в разделе «Неприкосновенность частной жизни / риски нарушения конфиденциальности».

### 3. Kumaй

Китай известен как страна, которая в первую очередь опирается на концепцию суверенитета данных и рассматривает данные как стратегический ресурс своей национальной экономики данных. Правовое регулирование национальной системы данных включает в себя три основных документа: Закон о кибербезопасности (Cybersecurity Law)<sup>2</sup>, Закон о защите данных (Data Security Law)<sup>3</sup> и Закон о защите личной информации (Personal Information Protection Law)<sup>4</sup>.

С точки зрения регулирования вопросов оборота данных наибольший интерес представляют ключевые положения Закона о защите данных. Обратимся к наиболее важным среди них:

– *категоризация данных*. Данные классифицируются исходя из их значения для национальной безопасности, публичного интереса и развития экономики, что позволяет правительству использовать различные охранительные режимы в зависимости от их чувствительности и потенциального влияния;

– *доступ государства к данным*. В рамках сохранения контроля над данными резидентов страны правительство обладает существенными полномочиями по доступу к этим данным, в том числе и обрабатываемым в частном секторе;

– *локализация данных и ограничение трансграничной передачи*. Вся чувствительная и критическая информация должны генерироваться и обрабатываться под юрисдикцией государства. Трансграничная передача допускается только при условии получения на то специального разрешения компетентного органа;

– *эффективное использование данных*. В противовес ограничениям внедряются инструменты поддержки развития технологий по обработке данных при условии, что они соответствуют установленным правилам.

Безусловно, данный обзор не дает общего представления о возможности использования синтетических данных, как и об их регулировании и применении в целях государственного управления, однако позволяет сформировать общее понимание существующих тенденций в правовом регулировании и их направленности в данной сфере.

Стоит отметить и то, что такое направление государственной политики демонстрирует создание единого подхода к обработке правительственных данных и их интеграции в единые озера данных. В частности, в 2022 году было анонсировано создание Интегрированной государственной системы больших данных (Integrated Government Big Data

□ technologies to avoid infringing on intellectual property rights;

□ generative AI cannot be utilized to create legally binding documents;

□ generative AI should not be utilized for critical and time-sensitive processes.

The EU Artificial Intelligence Act includes several critical regulations governing the use of AI in public administration, particularly concerning the special labeling of both the results generated by AI and instances of interaction between citizens and AI when accessing public services.

The Data Act also introduced significant changes to the regulation of data processing within the EU. In particular, the document grants authorities the right to access data collected and stored by private sector entities under specific conditions during public emergency situations.

Another significant document that actively shapes the development of the EU data market is the Open Data Directive<sup>1</sup>. This directive clearly establishes that it is necessary to disclose information collected by government authorities across six thematic categories: geospatial datasets; earth observation and environment; meteorological data; statistics; companies and company ownership data; mobility. The datasets must be made available free of charge, in machine-readable formats, and through APIs to facilitate their effective re-use. In general, the implementation of the Directive resulted in the emergence and development of the European Data Portal – a pan-European multilingual platform for searching public-sector information that is open for reuse.

In the considered strategy, in its part regarding the development of a data market in the EU, synthetic data is increasingly viewed as an alternative to anonymized data. This approach aims to mitigate the risk of infringing on the rights and interests of individuals when processing personal data [9], as elaborated in greater detail in the section titled ‘Privacy / Confidentiality Risks’.

### 3. China

China is widely recognized as a country that emphasizes the concept of data sovereignty, viewing data as a strategic resource within its national data economy. The legal framework governing the national data system comprises three main documents: the Cybersecurity Law<sup>2</sup>, the Data Security Law<sup>3</sup>, and the Personal Information Protection Law<sup>4</sup>.

From the perspective of regulating data circulation issues, the key provisions of data security law are of utmost importance. Below are listed the most significant of them:

□ *categorization of data*. Data is classified based on its importance for national security, public interest, and economic development, which allows the government to use various protective regimes depending on its sensitivity and potential impact;

□ *government access to data*. In order to retain control over data of the country’s residents, the government has significant powers to access this data, including data processed in the private sector);

□ *localization of data and restriction of cross-border transfer*. All sensitive and critical information must be generated and processed under the jurisdiction of the state. Cross-border transfer is allowed only if a special permit from the competent authority is obtained;

□ *efficient use of data*. As opposed to restrictions, tools are being introduced to support the development of data processing technologies, provided that they comply with the established rules).

Certainly, this brief presentation does not offer a comprehensive overview of the potential uses of synthetic data, nor does it address its regulation and application for public administration purposes. However, it does provide a general understanding of the current trends in their legal regulation.

It is also important to note that recent initiatives demonstrate the establishment of a unified approach to government data processing and its integration into consolidated data lakes. In particular, the announcement of the Integrated Government Big Data System

<sup>1</sup> Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information. URL: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2019.172.01.0056.01.ENG](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG).

<sup>2</sup> Cybersecurity Law of the People’s Republic of China (unofficial translation). URL: [https://www.dataguidance.com/sites/default/files/en\\_cybersecurity\\_law\\_of\\_the\\_peoples\\_republic\\_of\\_china\\_1.pdf](https://www.dataguidance.com/sites/default/files/en_cybersecurity_law_of_the_peoples_republic_of_china_1.pdf).

<sup>3</sup> Data Security Law of the People’s Republic of China. URL: [http://en.npc.gov.cn.cdurl.cn/2021-06/10/c\\_689311.htm](http://en.npc.gov.cn.cdurl.cn/2021-06/10/c_689311.htm).

<sup>4</sup> Personal Information Protection Law of the People’s Republic of China (Adopted at the 30th Meeting of the Standing Committee of the Thirteenth National People’s Congress on August 20, 2021). URL: [http://en.npc.gov.cn.cdurl.cn/2021-12/29/c\\_694559.htm](http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm).

<sup>1</sup> Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information. Available at: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2019.172.01.0056.01.ENG](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG)

<sup>2</sup> Cybersecurity Law of the People’s Republic of China (unofficial translation). *DataGuidance* [website]. Available at: [https://www.dataguidance.com/sites/default/files/en\\_cybersecurity\\_law\\_of\\_the\\_peoples\\_republic\\_of\\_china\\_1.pdf](https://www.dataguidance.com/sites/default/files/en_cybersecurity_law_of_the_peoples_republic_of_china_1.pdf).

<sup>3</sup> Data Security Law of the People’s Republic of China. *The National People’s Congress of the People’s Republic of China* [website]. Available at: [http://en.npc.gov.cn.cdurl.cn/2021-06/10/c\\_689311.htm](http://en.npc.gov.cn.cdurl.cn/2021-06/10/c_689311.htm)

<sup>4</sup> Personal Information Protection Law of the People’s Republic of China the National People’s Congress of the People’s Republic of China (adopted at the 30th Meeting of the Standing Committee of the Thirteenth National People’s Congress on August 20, 2021). Available at: [http://en.npc.gov.cn.cdurl.cn/2021-12/29/c\\_694559.htm](http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm)

System)<sup>1</sup>. Эта система включает три типа платформ: национальную платформу больших данных для государственных вопросов (главный центр управления данными для правительственных операций в Китае), платформу на уровне провинций (охватывает 31 провинцию, автономный район и муниципалитет) и платформу данных для государственных вопросов соответствующих департаментов Госсовета.

## Публичные озера данных (цифровые озера) в России

### 1. Цифровая аналитическая платформа

В 2019 году было отмечено, что многие используемые государственными органами инструменты для статистических исследований несовершенны, в связи с чем в методиках, форматах и показателях наблюдаются разночтения, которые ведут к разнице в интерпретациях. Эту проблему можно решить, переведя статистический учет в цифровой формат<sup>2</sup>.

Национальная программа «Цифровая экономика Российской Федерации» включает в себя федеральный проект «Цифровое государственное управление», предусматривающий создание национальной системы управления данными, одним из компонентов которой должна стать платформа данных. Предполагается, что платформа будет содержать такие элементы, как: национальная система управления данными; единый государственный реестр записей актов гражданского состояния; единый федеральный информационный ресурс, собирающий сведения о населении Российской Федерации; федеральная государственная информационная система «Единая система нормативной справочной информации».

Создание платформы будет осуществляться на базе Федеральной службы государственной статистики, которой предстоит выполнять функции ее оператора. В 2022 году вступило в силу соответствующее Постановление о государственной информационной системе «Цифровая аналитическая платформа предоставления статистических данных»<sup>3</sup>.

Система обеспечивает выполнение следующих функций:

- централизованный сбор, хранение, обработка и предоставление информации системы;

- формирование и ведение единого реестра объектов статистического наблюдения;

- формирование и ведение единого реестра форм статистического наблюдения и статистических показателей;

- формирование, ведение и мониторинг выполнения Федерального плана статистических работ;

- защита информации системы, обеспечение целостности и доступности обрабатываемой в ней информации;

- доступ к сведениям, содержащимся в системе, предоставление таких сведений в электронном виде, а также формирование аналитических отчетов на основе информации системы;

- информационное взаимодействие системы с информационными системами субъектов системы, пользователей системы, а также с иными информационными системами;

- разработка (корректировка), согласование и утверждение методик расчета показателей национальных проектов (программ), государственных программ Российской Федерации и их структурных элементов;

- расчет значений показателей национальных проектов (программ), государственных программ Российской Федерации и их структурных элементов, анализ достижения плановых значений указанных показателей.

Система является общедоступной, за исключением информации, доступ к которой ограничен законодательством Российской Федерации.

В 2024 году был принят Закон о цифровой платформе статистических данных<sup>4</sup>. Документ разработан в целях снижения отчетной нагрузки на субъекты малого предпринимательства, оптимизации форм и видов предоставляемой ими отчетности, а также предотвращения ее дублирования различными ведомствами.

### 2. Национальная система управления данными (НСУД)

Решением Правительства РФ в 2019 году были утверждены Концепция создания и функционирования национальной системы управления данными и План мероприятий («дорожная карта») по созданию национальной системы управления данными на 2019–2021 годы<sup>5</sup>.

<sup>1</sup> China to Build Integrated Govt Big Data System. URL: [https://english.www.gov.cn/policies/latestreleases/202210/28/content\\_WS635b9af9c6d0a757729e1e00.html](https://english.www.gov.cn/policies/latestreleases/202210/28/content_WS635b9af9c6d0a757729e1e00.html).

<sup>2</sup> Концепция создания цифровой аналитической платформы предоставления статистических данных: утв. распоряжением Правительства Рос. Федерации от 17 дек. 2019 г. № 3074-р. URL: <http://static.government.ru/media/files/4YejV8mvcCSeG-WTg2kXprmthtNbWyfrU.pdf>.

<sup>3</sup> О государственной информационной системе «Цифровая аналитическая платформа предоставления статистических данных»: Постановление Правительства Рос. Федерации от 22 июня 2021 г. № 956. URL: <http://government.ru/docs/all/135122/>.

<sup>4</sup> О внесении изменений в статью 5 Федерального закона «О развитии малого и среднего предпринимательства в Российской Федерации» и Федеральный закон «Об официальном статистическом учете и системе государственной статистики в Российской Федерации»: Федер. закон от 22 июля 2024 г. № 206-ФЗ (ред. от 23 мая 2025 г.) // Собр. законодательства Рос. Федерации. 2024. № 31, ст. 4466.

<sup>5</sup> Концепция создания и функционирования национальной системы управления данными и План мероприятий («дорожная карта») по созданию национальной системы управления данными на 2019–2021 годы: утв. распоряжением Правительства

in 2022<sup>1</sup> marked a significant development. This system comprises three types of platforms: the national big data platform for government affairs (the central data management center for government operations in China); the provincial-level platform (covering 31 provinces, autonomous regions, and municipalities), and the government affairs data platform for the relevant departments of the State Council.

## Public Data Lakes in Russia

### 1. National Analytical Platform

In 2019, it was observed that many tools utilized by government agencies for statistical research were flawed, this resulting in discrepancies in the methods, formats, and indicators, which, in turn, could lead to misinterpretation. It was suggested to address this issue by converting statistical accounting to a digital format<sup>2</sup>.

The national program ‘Digital Economy of the Russian Federation’ includes the federal project ‘Digital Public Administration’, which aims to establish a national data management system. One of the key components of this system is a data platform. It is anticipated that the platform will incorporate several elements, including the national data management system; the unified state register of civil status records; the unified federal information resource containing data about the population of the Russian Federation; and the federal state information system ‘Unified System of Normative Reference Information’.

The platform will be established based on the Federal State Statistics Service, which will act as its operator and integrator. In 2022, the relevant Regulation on the state information system ‘Digital Analytical Platform for Providing Statistical Data’ entered into force<sup>3</sup>.

The system is intended to perform the following functions:

- centralized collection, storage, processing, and provision of data;

- formation and maintenance of a unified register of statistical observation;

- formation and maintenance of a unified register of the forms of statistical observation and statistical indicators;

- formation, maintenance, and monitoring of the implementation of the Federal Statistical Works Plan;

- protection of information, ensuring of the integrity and accessibility of the information processed in the system;

- access to information stored in the system, provision of this information in electronic form, generation of analytical reports based on the system’s data;

- information interaction of the system with the information systems of its users, including state and municipal bodies, agencies, and with other information systems;

- development (correction), coordination, and approval of calculation methods for indicators of national projects (programs), state programs of the Russian Federation, and their structural elements;

- calculation of the values of indicators for national projects (programs) and state programs of the Russian Federation, along with their structural elements, analysis of the achievement of the values planned for these indicators.

The system is publicly accessible, except for information the access to which is restricted by the legislation of the Russian Federation.

In 2024, the digital statistical data platform was officially introduced by law<sup>4</sup>. The document was developed to alleviate the reporting burden on small businesses, optimize the forms and types of data reporting required from them, and prevent the duplication of requests made by various public agencies and bodies.

### 2. National Data Management System (NDMS)

The Government of the Russian Federation adopted the Concept of the Creation and Operation of the National Data Management System, along with the Action Plan (Roadmap) for the Creation of the National Data Management System for 2019–2021, in 2019 (hereinafter referred to as the Concept)<sup>5</sup>.

<sup>1</sup> China to build integrated govt big data system. *The State Council of Public Republic of China* [website]. Available at: [https://english.www.gov.cn/policies/latestreleases/202210/28/content\\_WS635b9af9c6d0a757729e1e00.html](https://english.www.gov.cn/policies/latestreleases/202210/28/content_WS635b9af9c6d0a757729e1e00.html)

<sup>2</sup> The government has approved the digitalization of statistics collection for an objective picture of the economy. CNews. Available at: [https://www.cnews.ru/news/top/2019-12-23\\_vlasti\\_utverdili\\_kontseptsiyu?ysclid=m7qvw2wdaa222415517](https://www.cnews.ru/news/top/2019-12-23_vlasti_utverdili_kontseptsiyu?ysclid=m7qvw2wdaa222415517). [В русской версии разнотой, выделен, возможно требует проверки Shmyrova V. The government has approved the digitalization of statistics collection for an objective picture of the economy. Available at: [https://www.cnews.ru/news/top/2019-12-23\\_vlasti\\_utverdili\\_kontseptsiyu](https://www.cnews.ru/news/top/2019-12-23_vlasti_utverdili_kontseptsiyu). ??? or the Concept of creating a digital analytical platform for the provision of statistical data: Approved by the decree of the Government of the Russian Federation No. 3074-r of December 17, 2019. Available at: <http://static.government.ru/media/files/4YejV8mvcCSeG-WTg2kXprmthtNbWyfrU.pdf>.]

<sup>3</sup> ‘On the State Information System ‘Digital Analytical Platform for Providing Statistical Data’: Decree of the Government of the Russian Federation No. 956 of June 22, 2021. Available at: <http://government.ru/docs/all/135122/>.

<sup>4</sup> ‘On the Introduction of Amendments to Article 5 of the Federal Law ‘On the Development of Small and Medium-Sized Entrepreneurship in the Russian Federation’ and the Federal Law ‘On Official Statistical Accounting and the State Statistics System in the Russian Federation’: Federal Law No. 206-FZ of July 22, 2024 (as of May 23, 2025). *Collection of Legislative Acts of the Russian Federation*. 2024. No. 31. Art. 4466.

<sup>5</sup> The Concept of the Creation and Operation of the National Data Management System and the Action Plan (Roadmap) for the Creation of the National Data Management System for 2019–2021: approved by Order of the Government of the Russian Federation No. 1189-r of June 3, 2019 (as of May 14, 2021). Available at: <http://static.government.ru/media/files/jYh27VlwiZs44qa0IXJZCa3uu7qqLzL.pdf>

Под национальной системой управления данными понимается система, состоящая из взаимосвязанных элементов информационно-технологического, организационного, методологического, кадрового и нормативно-правового характера и обеспечивающая достижение целей и выполнение задач, обозначенных в Концепции. При этом государственными данными являются сведения, содержащиеся в информационных ресурсах органов и организаций государственного сектора, а также в информационных ресурсах, созданных в целях реализации полномочий органов и организаций государственного сектора.

Положения Концепции также распространяются на информацию, не являющуюся государственными данными, при ее предоставлении для включения в информационные ресурсы органов и организаций государственного сектора, а также при ее участии в информационном обмене между органами и организациями государственного сектора с учетом требований законодательства Российской Федерации в части соблюдения прав и законных интересов владельцев данных, не являющихся государственными данными, а также соблюдения принципов невмешательства в их хозяйственную деятельность.

В систему входят следующие взаимосвязанные элементы:

1) совокупность нормативных правовых, организационных, методологических правил и процедур, регулирующих отношения органов и организаций государственного сектора, юридических и физических лиц в сфере управления государственными данными, а также обеспечивающих деятельность участников системы;

2) федеральная государственная информационная система «Единая информационная платформа национальной системы управления данными» и иные информационно-технологические элементы системы;

3) цифровая аналитическая платформа предоставления статистических данных.

Обработка сведений, составляющих государственную тайну, в единой информационной платформе не допускается. Исчерпывающий перечень сведений ограниченного доступа, доступ к которым предоставляется посредством единой информационной платформы, определяется в соответствии с законодательством Российской Федерации.

В 2021 году Постановлением Правительства РФ утверждено Положение о федеральной государственной информационной системе «Единая информационная платформа национальной системы управления

данными» (далее – НСУД)<sup>1</sup>. НСУД «входит в инфраструктуру, обеспечивающую информационно-технологическое взаимодействие информационных систем, используемых для предоставления государственных и муниципальных услуг и исполнения государственных и муниципальных функций в электронной форме, и создается в целях повышения эффективности обмена и использования государственных данных для предоставления государственных и муниципальных услуг и исполнения государственных и муниципальных функций в электронной форме и иных данных» (п. 3).

НСУД предназначена для решения следующих задач:

- систематизация государственных данных;
- формирование и ведение моделей государственных данных;
- формирование и ведение реестра информационных ресурсов;
- автоматизация процессов управления государственными данными;
- обеспечение контроля качества государственных данных;
- обеспечение сбора, обработки, хранения, предоставления, визуализации и анализа государственных данных из информационных систем органов и организаций государственного сектора и иных данных;
- формирование матрицы доступа.

Участниками НСУД являются органы и организации государственного сектора, а также иные организации, чьи информационные системы участвуют в информационном обмене посредством системы взаимодействия. Органы и организации государственного сектора выступают в роли обладателей государственных данных и пользователей государственных данных. Иные организации, использующие систему взаимодействия, выступают в качестве пользователей государственных данных.

Важно отметить, что уже сейчас НСУД предоставляет по запросу наборы данных для искусственного интеллекта (распознавание такси на изображениях, автотранспортных средств на основе данных БПЛА-съемки; детекция износа и повреждений поверхности дорожного покрытия; распознавание различных транспортных средств для целей определения выбросов загрязняющих веществ в атмосферный воздух от передвижных источников для проведения сводных расчетов загрязнения атмосферного воздуха; распознавание общественного транспорта на изображениях (автобусы, трамваи, троллейбусы) и многое другое).

The national data management system (hereinafter also referred to as the NDMS) is defined as a framework comprising interrelated components of an information-technology, organizational, methodological, personnel-related, and regulatory nature and ensuring the achievement of the goals and objectives outlined in the Concept. Government data is understood to mean information contained in the information resources of public sector bodies and organizations as well as in the information resources created to facilitate performance by these entities of their functions.

The provisions of the Concept also apply to information that is not inherently government data when it is specifically designated for inclusion in the information resources of public sector bodies and organizations. This includes instances where the information is part of an exchange between public sector entities, ensuring the protection of the rights and legitimate interests of its original owners – private individuals or companies – while also safeguarding against interference in their economic activities.

The system comprises the following interrelated elements:

1) a set of regulatory, legal, organizational, and methodological rules and procedures governing the relationships among the public sector bodies, organizations, legal entities, and individuals in the realm of government data management, and also ensuring the effective functioning of the system's participants.

2) the federal state information system 'Unified Information Platform of the National Data Management System' and other information-technology components of the system;

3) a digital analytical platform for delivering statistical data.

The system cannot be used for processing state secrets. An exhaustive list of restricted access information, which is stored and provided by the platform, is determined in accordance with the legislation of the Russian Federation.

In 2021, the government adopted the Regulation on the federal state information system 'Unified Information Platform of the National Data Management

System'<sup>1</sup>. According to its provisions, the NDMS is an integral part of the infrastructure that facilitates the information and technological interaction of information systems used to deliver state and municipal services and to perform state and municipal functions in electronic form; it is created to enhance the efficiency of state data exchange and utilization for the provision of state and municipal services and the execution of state and municipal functions electronically (Item 3).

The NDMS is intended to perform the following tasks:

- systematization of government data;
- formation and maintenance of government data models;
- formation and maintenance of a register of information resources;
- automation of government data management processes;
- ensuring the quality control of government data;
- ensuring the collection, processing, storage, provision, visualization, and analysis of government data obtained from the information systems of public sector bodies, organizations, and other data sources;
- formation of an access matrix.

The NDMS participants include public sector bodies and organizations as well as other entities whose information systems are engaged in information exchange through the interaction system. The public sector bodies and organizations serve as both holders and users of government data, while other organizations utilizing the NDMS function solely as users of state data.

Importantly, it is already the case that the NDMS can provide datasets for AI processing upon request. These datasets are used for the recognition of taxis in images, identification of vehicles based on UAV survey data, detection of road wear and damage to road surfaces, recognition of various vehicles for assessing emissions of pollutants from mobile sources for atmospheric air pollution calculations, identification of public transport (buses, trams, trolleybuses) in images, and for many other applications.

Рос. Федерации от 3 июня 2019 г. № 1189-р (ред. от 14 мая 2021 г.). URL: <http://static.government.ru/media/files/jYh27VtwiZs44qa0IXJZCa3uu7qqLzl.pdf>.

<sup>1</sup> Об утверждении Положения о федеральной государственной информационной системе «Единая информационная платформа национальной системы управления данными»: Постановление Правительства Рос. Федерации от 14 мая 2021 г. № 733 (в ред. от 28 ноября 2024 г.). URL: <http://static.government.ru/media/files/UN7KKWcK3TgpjAmRcaTzRWUZezbiQQWf.pdf>.

<sup>1</sup> 'On Approval of the Regulation on the Federal State Information System 'The Unified Information Platform of the National Data Management System': Decree of the Government of the Russian Federation No. 733 of May 14, 2021 (as of November 28, 2024). Available at: <http://static.government.ru/media/files/UN7KKWcK3TgpjAmRcaTzRWUZezbiQQWf.pdf>

### 3. Цифровое озеро обезличенных персональных данных

В 2024 году Федеральным законом от 8 августа 2024 г. № 233-ФЗ были внесены изменения в Федеральный закон «О персональных данных» и Федеральный закон «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона “О персональных данных”»<sup>1</sup>. Закон направлен на обеспечение благоприятных правовых условий для сбора, хранения и обработки данных с использованием новых технологий и вступил в силу 1 сентября 2025 г. Принятие закона призвано существенно повысить эффективность системы защиты прав субъектов персональных данных, а также предоставить им право на использование методологической базы по обезличиванию персональных данных и их передачу в специальную государственную информационную систему (ГИС) для последующего использования.

Закон о персональных данных<sup>2</sup> дополнен новыми положениями, предусматривающими предоставление операторами персональных данных (ПДн) по требованию Минцифры обезличенных персональных данных в государственную информационную систему министерства, которую определит Правительство РФ. В соответствующем требовании будет указываться перечень данных, которые необходимо передать, а также сроки их предоставления. Минцифры будет формировать из обезличенных данных, сгруппированных по определенному признаку, наборы персональных данных.

При этом установлено, что:

– формирование наборов данных будет осуществляться Минцифры исключительно в случаях, которые определит Правительство РФ;

– формирование наборов данных из специальных категорий персональных данных (есть исключение) и биометрических персональных данных не допускается;

– доступ к дата-сетам будет происходить исключительно в рамках закрытого контура государственной

информационной системы, «вынести» их из нее будет невозможно;

– доступ к сформированным дата-сетам пользователей ГИС – государственных и муниципальных органов и подведомственных им организаций, органов государственных внебюджетных фондов; граждан РФ и российских юридических лиц, сведения о которых внесены в реестр операторов персональных данных и которые соответствуют ряду требований, – будет иметь определенные особенности;

– предоставление результатов обработки составов данных иностранным юридическим лицам, иностранным организациям, не являющимся юридическими лицами, иностранным гражданам и лицам без гражданства по общему правилу запрещается.

Департамент информационных технологий города Москвы будет наделен аналогичными правами по созданию регионального (локального) хранилища обезличенных данных, полученных из государственных и муниципальных информационных систем. Минцифры, в свою очередь, будет иметь право запрашивать обезличенные данные из регионального хранилища (озера данных)<sup>3</sup>. Закон предусматривает, и это важно отметить, что набор данных, полученных в результате обезличивания, формируется при условии, что последующая обработка этих данных не позволит установить их принадлежность конкретному физическому лицу<sup>4</sup>.

#### Использование синтетических данных в целях государственного управления

##### 1. Неприкосновенность частной жизни / риски нарушения конфиденциальности

На сегодняшний день этот вопрос является чрезвычайно важным как с точки зрения правового регулирования, так и с точки зрения практической, в части выбора соответствующих технологий и алгоритмов, которые бы обеспечили необходимый уровень анонимности, то есть деперсонализации данных. Современное законодательство о персональных данных называет таковыми любую информацию, которая может быть отнесена к «определенному или определяемому на основании такой информации индивиду»<sup>5</sup>.

<sup>1</sup> О внесении изменений в Федеральный закон «О персональных данных» и Федеральный закон «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона “О персональных данных”»: Федер. закон от 8 авг. 2024 г. № 233-ФЗ // Собр. законодательства Рос. Федерации. 2024. № 33 (ч. 1), ст. 4929.

<sup>2</sup> О персональных данных: Федер. закон от 27 июля 2006 г. № 152-ФЗ (в ред. от 28 февраля 2025 г.) // Собр. законодательства Рос. Федерации. 2006. № 31 (ч. 1), ст. 3451.

<sup>3</sup> С 1 сентября 2025 года компании начнут передавать обезличенные сведения в ГИС. URL: <https://www.garant.ru/news/1749290/>.

<sup>4</sup> Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. URL: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2019.172.01.0056.01.ENG](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG).

<sup>5</sup> Например: О персональных данных: Федер. закон от 27 июля 2006 г. № 152-ФЗ; General Data Protection Regulation. URL: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legisum:310401\\_2](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legisum:310401_2).

### 3. National Data Lake of Depersonalized Personal Data

In 2024, Federal Law No. 233-FZ<sup>1</sup> amended the Federal Laws ‘On Personal Data’ and ‘On the Conduct of an Experiment on Establishing Special Regulation in Order to Create the Necessary Conditions for the Development and Introduction of Artificial Intelligence Technology in the Constituent Entity of the Russian Federation – the Federal City of Moscow and on the Introduction of Amendments to Articles 6 and 10 of the Federal Law On Personal Data’. The law aims to create favorable legal conditions for the collection, storage, and processing of data using new technologies and came into effect on September 1, 2025. The adoption of this law was intended to significantly enhance the effectiveness of the system designed to protect the rights of individuals whose personal data is processed. The law grants individuals the right to utilize the methodological framework for the depersonalization of their personal data and to transfer this data to a specialized state information system (SIS) for subsequent use (re-use).

The Law ‘On Personal Data’<sup>2</sup> has been amended to include new provisions requiring that data processors transmit depersonalized data upon the request of the Ministry of Digital Development, Communications, and Mass Media (hereinafter referred to as the Ministry) to the Ministry’s state information system, which will be designated by the Government of the Russian Federation. The relevant request is supposed to state the specific types of data to be transmitted as well as the timeframe for the provision of data. The Ministry will compile sets of personal data from depersonalized data categorized by specific attributes.

It is established that:

□ the datasets will be generated exclusively by the Ministry in cases determined by the Government of the Russian Federation;

□ the creation of datasets from special (sensitive) categories of personal data (with an exception specified) and from biometric personal data is prohibited;

□ access to datasets will occur exclusively within the closed loop of the state information system, and it will be impossible to remove them from this system;

□ access to the datasets for the SIS users – state and municipal bodies, organizations subordinate to them, bodies of state extra-budgetary funds, Russian citizens and Russian legal entities that are included in the register of personal data processors and meet a number of requirements – will have certain features;

□ it is prohibited, as a general rule, to disclose the results of processing datasets to foreign legal entities, foreign organizations that are not legal persons, foreign citizens, and stateless persons.

The Department of Information Technologies of the City of Moscow will have similar rights to establish regional (local) data storage for depersonalized data obtained from state and municipal information systems. The Ministry, in turn, will have the authority to request the depersonalized data from the regional storage (data lake)<sup>3</sup>. Importantly, the law stipulates that a dataset obtained through depersonalization is created under the condition that subsequent processing of this data will not enable the identification of whether it belongs to a specific individual<sup>4</sup>.

#### The Use of Synthetic Data for Public Administration Purposes

##### 1. Privacy / Confidentiality Risks

This issue is highly important from both legal and practical perspectives, particularly in terms of the selection of appropriate technologies and algorithms that would ensure a necessary level of anonymity, depersonalization of data. Modern legislation on personal data refers to this any information that can be linked to an individual that is identified or identifiable based on that information<sup>5</sup>.

<sup>1</sup> ‘On the Introduction of Amendments to the Federal Law ‘On Personal Data’ and the Federal Law ‘On the Conduct of an Experiment on Establishing Special Regulation in Order to Create the Necessary Conditions for the Development and Implementation of Artificial Intelligence Technology in the Constituent Entity of the Russian Federation – the Federal City of Moscow and on the Introduction of Amendments to Articles 6 and 10 of the Federal Law ‘On Personal Data’: Federal Law No. 233-FZ of August 8, 2024. *Collection of Legislative Acts of the Russian Federation*. 2024. No. 33 (Pt. 1). Art. 4929.

<sup>2</sup> ‘On Personal Data’: Federal Law No. 152-FZ of July 27, 2006 (as of February 28, 2025). *Collection of Legislative Acts of the Russian Federation*. 2006. No. 31 (Pt. 1). Art. 3451.

<sup>3</sup> From September 1, 2025, companies will begin transferring depersonalized data to the State Information System. Available at: <https://www.garant.ru/news/1749290/>

<sup>4</sup> Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information. Available at: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2019.172.01.0056.01.ENG](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG)

<sup>5</sup> For example: ‘On Personal Data’: Federal Law No. 152-FZ of July 27, 2006; General Data Protection Regulation. Available at: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legisum:310401\\_2](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legisum:310401_2)

Отсюда возникает вполне закономерный вопрос об алгоритмах и технологиях преобразования реальных данных при создании из них синтетических. Сама технология производства синтетических данных подразумевает как относительно «легкое» преобразование реальных данных (а следовательно, такая информация может содержать отдельные элементы синтетических данных или в дальнейшем приводить к ситуации, когда риск соотнесения ее с тем или иным индивидом объективно сохраняется), так и «полную» их трансформацию (когда полностью синтетические данные повторяют лишь общие шаблоны, рисунки реальных данных) [5].

Эта проблематика относительно давно известна исследователям в области приватности в европейских странах, где также сформировались два процесса преобразования персональных данных для снижения риска их последующего использования: псевдонимизация и анонимизация<sup>1</sup>. Первая предполагает в целом сохранение реальных данных, заменяя часть из них псевдонимами и (или) идентификаторами, и допускает обратное преобразование; вторая трансформирует данные более глубоко, в результате чего в идеале утрачивается связь между набором данных и индивидом. Отчасти это напоминает создание синтетических данных. «Легкие» синтетические данные по своему правовому режиму вполне могут рассматриваться как данные, подвергнутые псевдонимизации, и требовать соблюдения общих правил для защиты персональных данных, пусть и на относительно низком уровне с учетом потенциальных рисков для прав индивида. Полностью синтетические данные, с другой стороны, вполне можно рассматривать как анонимизированные данные, то есть как способ полной анонимизации данных [9].

Такая проблематика активно обсуждается и в России, где, как и в других странах, существует реальная потребность бизнеса в получении доступа к большим массивам данных и в выработке условий, при которых данные утрачивают признаки персональных, то есть становятся анонимными. Однако российское законодательство в этой сфере противоречиво, поскольку использует только один вариант описания таких действий – «обезличивание», которое в нормативных документах представляет собой аналог псевдонимизации, или относительно легкую модификацию данных, сохраняющую возможность их обратного преобразования. Этой же позиции придерживается российский регулятор Роском-

надзор, представители которого считают, что обезличенные персональные данные являются по-прежнему персональными данными с сопутствующими рисками необходимости их защиты [9; 2]. Российская практика и исследователи за последнее время неоднократно обосновывали необходимость разработки алгоритмов (правил) гарантированного обезличивания, которое, по сути, является аналогом анонимизации, то есть такого преобразования данных, которое исключает возможность идентификации индивида или установления связи между ним и данными [3].

Аналогичная ситуация и с иной конфиденциальной информацией в части сохранения рисков раскрытия чувствительной информации при использовании «легких» синтетических данных.

## 2. Риски принятия решений

Использование синтетических данных несет в себе определенные риски в принятии решений на основе указанных данных. Учеными давно отмечается, что сфера государственного управления относительно консервативна к инновациям и требует тщательного выбора средств для своей реализации. Это в особенности касается выбора технологий искусственного интеллекта. В данном случае стоит признать справедливым мнение сторонников установления четкой и последовательной политики в части ограничений на использование и применение технологий генеративного искусственного интеллекта в области публичного управления [7].

С учетом этого вполне очевидной становится проблема проведения испытаний (тестирования), а также оценки применимости технологий искусственного интеллекта для целей публичного управления, вплоть до проведения оценки результатов и алгоритмов искусственного интеллекта экспертами, обладающими необходимыми навыками и компетенциями [18].

Именно в таком случае синтетические данные имеют массу преимуществ, поскольку несут меньшие риски для раскрытия чувствительной информации, но при этом позволяют получить результаты работы алгоритмов генеративного искусственного интеллекта и оценить перспективы и возможности его использования для целей публичного управления. Таким образом, можно привлечь к разработке или оценке искусственного интеллекта частные компании без риска раскрытия конфиденциальной информации или нанесения иного вреда деятельности по публичному управлению.

This raises a pertinent question regarding the algorithms and technologies used to convert real data into synthetic data. The process of producing synthetic data involves both relatively ‘light’ transformation (hybrid or semisynthetic) and more comprehensive transformations, resulting in ‘fully’ synthetic data that only replicates the general patterns of real data. Consequently, even hybrid or semisynthetic data may contain elements that correlate with specific individuals, thereby posing a risk of re-identification [5].

In this context, the issue has long been recognized by privacy researchers in European countries, where developed two distinct processes for transforming personal data to mitigate the risk of subsequent misuse: pseudonymization and anonymization<sup>1</sup>. The first approach typically involves retaining real data while substituting some of it with aliases and/or identifiers, allowing for the possibility of reidentification. In contrast, the second method entails a more profound transformation of data, ideally resulting in the complete severance of the connection between the dataset and the individual. This is somewhat similar to the creation of synthetic data. In terms of its legal framework, ‘light’ synthetic data may well be regarded as pseudonymized data and thus necessitate adherence to general regulations for the protection of personal data, albeit at a relatively low level, considering the potential risks to individual rights. Conversely, ‘fully’ synthetic data can be regarded as anonymized data, i.e., as a method of complete data anonymization [9].

This issue is currently being actively discussed in Russia, where businesses demonstrate a significant demand for access to large volumes of data. Additionally, there is a need to establish conditions under which data loses its characteristics, meaning it becomes anonymous. However, Russian legislation in this area is relatively contradictory: it employs only one term to describe such actions – ‘depersonalization’, while in regulatory documents this term serves as an analogue to pseudonymization or a relatively simple modification of data that retains the possibility of reverse transformation. The Russian regulator, Roskomnadzor, has consistently

maintained that ‘depersonalized’ personal data is still considered personal data, which carries associated risks that necessitate protection [9], [2]. Recently, Russian practitioners and researchers have increasingly emphasized the need to develop algorithms and rules for guaranteed depersonalization. This process is essentially an analogue to anonymization, i.e., transformation of data that eliminates the possibility of identifying an individual or establishing a connection between the individual and the data [3].

The situation is comparable with other confidential information within a similar framework regarding the risks of disclosing sensitive information when utilizing ‘light’ synthetic data.

## 2. Decision-Making Risks

The use of synthetic data carries certain risks where decisions are made based on it. Researchers have repeatedly noted that the field of public administration is relatively conservative regarding innovation and requires careful selection of the methods employed for its implementation. This is particularly true for the choice of artificial intelligence technologies. In this context, it is important to acknowledge the perspective of those advocating for the establishment of a clear and consistent policy concerning restrictions on the use and application of generative AI technologies in public administration [7].

Given that, the challenges of conducting trials and testing, as well as assessing the applicability of AI technologies for public administration purposes, become evident. This includes the evaluation of AI results and algorithms by experts possessing the necessary skills and competencies [18].

It is in this context that synthetic data offers numerous advantages as it poses lower risks of disclosing sensitive information. At the same time, it enables us to obtain results from the algorithms of generative AI and assess them to determine their potential applications in public administration. Consequently, private companies can participate in the development or evaluation of AI without the risk of revealing confidential information or adversely affecting public administration activities.

<sup>1</sup> Pseudonymization according to the GDPR [definitions and examples] // Data Privacy Manager. 2021. Nov. 2. URL: <https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/>.

<sup>1</sup> Pseudonymization according to the GDPR [definitions and examples]. *Data Privacy Manager* [website]. 2021. 2 November. Available at: <https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/>

Однако следует понимать, что и при использовании синтетических данных присутствуют риски ошибочного принятия решений. Основной проблемой является дилемма между степенью преобразования данных и, как следствие, их точностью, относимостью и рисками раскрытия чувствительной информации. Слишком глубокое и полное преобразование данных при получении полностью синтетических данных несет риски искажения данных, в том числе паттернов данных. В итоге это может отразиться и на процессах принятия решений, которые будут реализовываться с использованием технологий искусственного интеллекта, обученных или построенных на основе таких искаженных искусственных данных. Единственным решением видится достижение определенного баланса между степенью раскрытия чувствительной информации и целями разработки качественных и надежных решений искусственного интеллекта для целей публичного управления, а также проведение тщательной оценки применяемых генеративным искусственным интеллектом синтетических данных [11].

### **3. Риски технологического суверенитета / риски суверенитета данных**

При выстраивании правового регулирования применения синтетических данных важным аспектом могут стать и вопросы обеспечения технологического суверенитета и суверенитета данных. Основная идея создания синтетических данных – обеспечение современной экономики данными важным и необходимым ресурсом. С этой стороны синтетические данные являются ответом на проблематику доступа к чувствительным данным, позволяя получить данные со сходными характеристиками без риска нанесения ущерба правам личности или государства.

Однако современное политическое противостояние государств на международной арене все чаще ставит вопросы обеспечения безопасности и сохранения технологического лидерства, равно как и необходимости основываться на собственных разработках. С учетом этого формируемые государством озера данных и в потенциале полученные на их основе синтетические данные становятся стратегическим ресурсом национальной экономики, доступ к которому требует тщательного и взвешенного подхода.

В сложившейся ситуации стоит обратить внимание на формирующийся подход в отношении предоставления доступа к цифровому озеру обезличенных персональных данных, о котором говорилось в одноименном разделе выше. Согласно правилам, доступ к нему предполагается только для аккредитованных компаний, то есть для российских

компаний, которым государство может доверить работу с такими данными, ожидая определенных позитивных результатов для экономики. Видится вполне объяснимым расширение указанного подхода при обеспечении доступа или использовании национальных озер данных для создания на их основе синтетических данных, а также их последующего реиспользования. Более того, современные российские исследования демонстрируют, что определенную опасность может нести использование данных о характеристиках социальных групп, для которых был предложен термин «данные о человеке». По сути, эта тенденция имеет рациональное обоснование, поскольку данные о характеристиках социальных групп могут быть потенциально использованы для осуществления недружественных действий в отношении государства и (или) указанных социальных групп, что предполагает создание соответствующего охранительного режима.

Синтетические данные с учетом их потенциальных свойств вполне могут характеризовать паттерны социальных групп и категорий индивидов, а также ситуацию в тех или иных отраслях экономики, что, безусловно, ставит вопрос если не об ограничительном режиме, то о необходимости управления и контроля за доступом к таким данным во избежание неблагоприятных последствий для государства [4].

#### **Заключение.**

##### **Перспективные направления развития**

Цифровое озеро может быть источником информации для обучения генеративного искусственного интеллекта. При этом цифровые озера могут создаваться как коммерческими организациями, так и государственными структурами, которые могут наполнять эти базы данных определенной информацией. Правовой режим использования искусственным интеллектом информации из цифрового озера может быть установлен владельцем цифрового озера либо соответствующим нормативным правовым актом (национальное законодательство), в том числе путем применения экспериментальных правовых режимов в сфере цифровых инноваций.

В этом случае важнейшими условиями получения доступа к такой информации будут являться:

- обеспечение защиты прав и свобод человека и гражданина и защита их персональных данных;
- обеспечение конфиденциальности полученных сведений;
- создание особого доступа к информации иностранных субъектов;
- наличие возможности быстрого прекращения доступа к информации.

However, it is important to recognize the risks associated with erroneous decision-making when using synthetic data. The primary challenge lies in balancing data transformation (to what extent data is transformed) with its accuracy, relevance, and risks of disclosing sensitive information. Excessive and comprehensive data transformation, which results in fully synthetic data, can lead to data distortion, including the alteration of underlying data patterns. Consequently, the use of such distorted synthetic data may adversely impact decision-making processes that rely on AI technologies trained or developed using this compromised data. The optimal approach appears to be finding a balance between the need to protect sensitive information and the objectives of creating high-quality, reliable AI solutions for public administration. Additionally, it is crucial to conduct a thorough evaluation of the synthetic data generators employed by AI [11].

### **3. Technological Sovereignty / Data Sovereignty Risks**

When developing legal regulations for the use of synthetic data, considerations regarding technological sovereignty and data sovereignty may also emerge as significant factors. The primary objective of synthetic data is to furnish the modern data economy with a vital and essential resource. In this context, synthetic data serves as a solution to the challenge of accessing sensitive information, enabling the generation of data with comparable characteristics while mitigating the risk of infringing upon individual or state rights.

However, the contemporary political confrontations among states in the international arena increasingly raise questions regarding the assurance of security and the maintenance of technological leadership, as well as the necessity to build upon their own advancements. In this context, the data lakes established by the state and the synthetic data potentially derived from them are emerging as strategic resources for the national economy, access to which necessitates a careful and balanced approach.

In this context, it is important to consider the emerging approach to providing access to the national lake of depersonalized data, which was discussed in the relevant section above. According to the regulations,

access is granted exclusively to accredited companies, specifically Russian firms that the government deems trustworthy to handle such data, with the expectation of achieving certain positive outcomes for the economy. It seems quite reasonable to expand this approach by utilizing national data lakes to create synthetic data and facilitate its subsequent reuse. Furthermore, contemporary Russian research indicates that using data on the characteristics of social groups, referred to as ‘human data’, could bear certain risks. Such data could potentially be exploited to conduct hostile actions against the state and (or) these social groups, which necessitates an appropriate protective regime to be established.

Synthetic data, given its potential properties, can effectively characterize the patterns of social groups and categories of individuals as well as the conditions in specific sectors of the economy. This raises the question of, if not a restrictive regime, then at least the need to manage and control access to such data in order to prevent adverse consequences for the state [4].

#### **Conclusion. Further Development**

A digital lake can serve as a valuable source of information for training generative artificial intelligence. Simultaneously, digital lakes can be established by both commercial organizations and government entities, which can enrich these databases with specific information. The legal framework governing the use of information from a digital lake by artificial intelligence can be determined by the owner of the digital lake or through relevant regulatory legal acts (national legislation), including the implementation of experimental legal regimes in the field of digital innovation.

In this case, the most important conditions for gaining access to such information are as follows:

- ensuring the protection of human and civil rights and freedoms as well as safeguarding personal data;
- ensuring the confidentiality of the information received;
- special access to information from foreign entities;
- the ability to quickly terminate or block access to information.

Цифровое озеро может заполняться синтетическими данными, когда информация в цифровое озеро поступает от генеративного искусственного интеллекта, который создает новую или преобразует имеющуюся информацию. Важно предусмотреть способы и технические средства, позволяющие идентифицировать реальную и синтетическую информацию, содержащуюся в цифровом озере.

Цифровое озеро может содержать информацию, имеющую побочный характер, – например, изображение автобуса (исходная информация), в котором находятся люди, или человека, переходящего дорогу перед автобусом (побочная информация). Таким образом, искусственный интеллект получает доступ не только к разрешенной информации, но и к сопутствующей информации, что является не совсем корректным.

Наконец, цифровое озеро может быть создано полностью из синтетических данных.

На основании изложенного можно также сделать определенные выводы о перспективах использования формируемых в России озер данных с целью создания синтетических данных для нужд публичного управления и правового регулирования в этой сфере:

1. Использование синтетических данных открывает перспективы для поиска, разработки и оценки применимости технологий искусственного интеллекта в сфере публичного управления. С учетом общей тенденции на повышение «вооруженности» государства технологиями генеративного искусственного интеллекта, обеспечивающими его эффективность, синтетические данные на основе государственных данных способны стать относительно безопасной базой для разработки технологических решений и научных исследований в этой сфере.

2. Правовой режим синтетических данных требует дальнейшего исследования и разработки правовых конструкций. Необходимо как минимум:

- разграничение/соотнесение режимов псевдонимизации, анонимизации и обезличивания (гарантированного обезличивания) данных с алгоритмами и инструментами создания синтетических данных;
- определение параметров охранительного режима синтетических данных, когда их использование может нести риски нанесения ущерба личности или государству;
- разграничение доступа к национальным озерам данных для создания синтетических данных и правил последующего реиспользования синтетических данных.

3. Важным аспектом внедрения синтетических данных в систему публичного управления должна стать процедура и правила оценки и применимости алгоритмов создания синтетических данных для нужд публичного управления в части их точности и соответствия заявленным целям и исключения возможных ошибок и противоречий.

#### Библиографический список

1. *Филющенко Л. И.* Цифровой профиль работника: проблемы защиты персональных данных // *Законы России: опыт, анализ, практика.* 2024. № 11. С. 93–97.
2. *Кинякина Е.* Обезличенные данные приравнивают к персональным // *Ведомости.* 2021. 12 марта. URL: <https://www.vedomosti.ru/technology/articles/2021/03/11/861164-obezhichennye-dannie>.
3. *Сазонова М.* Обезличивание данных: сохранение баланса между правами граждан и развитием инноваций // *Гарант.ру.* 2021. 24 мая. URL: <https://www.garant.ru/news/1464529/>.
4. *Соловкин С. В.* Публично-правовое регулирование процессов автоматизированного сбора данных о человеке в контексте цифровой трансформации: дис. ... канд. юрид. наук. М., 2025. 278 с.
5. *Abramov M.* Synthetic Data Definition: Pros and Cons. URL: <https://keymakr.com/blog/synthetic-data-definition-pros-and-cons/>.
6. *Bullman H.* The DoD and U.S. Census Bureau Turn to Data Lakes for Modernization and Greater Insights // *Government Technology Insider.* 2020. Aug. 6. URL: <https://governmenttechnologyinsider.com/the-dod-and-u-s-census-bureau-turn-to-data-lakes-for-modernization-and-greater-insights/>.
7. *Bundin M., Martynov A., Aliev Y., Kutuev E.* Legal Aspects of the Use of AI of Public Sector // *Digital Transformation and Global Society (Series: Communications in Computer and Information Science).* Vol. 858. Springer Cham, 2018. Pp. 171–180. DOI: 10.1007/978-3-030-02843-5\_14.
8. *Dixon J.* Pentaho, Hadoop, and Data Lakes. URL: <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes>.
9. *Farrell E. et al.* European Data Spaces – Scientific Insights into Data Sharing and Utilisation at Scale. Luxembourg: Publications Office of the European Union, 2023. 78 p. DOI: 10.2760/400188.
10. *Inmon W. H.* Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump. N.J.: Technics Publications, 2016. 166 p.
11. *Jälmy M., Elvander F., Waterschoot T.* Multi-Channel Low-Rank Convolution of Jointly Compressed Room Impulse Responses // *IEEE Open Journal of Signal Processing.* 2024. Vol. 5. Pp. 850–857. DOI: 10.1109/OJSP.2024.3410089.
12. *Kitchin R.* The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. London: SAGE Publications Ltd., 2014. 240 p. DOI: 10.4135/9781473909472.
13. *Mergel I., Edelmann N., Haug N.* Defining Digital Transformation: Results from Expert Interviews // *Government Information Quarterly.* 2019. Vol. 36. Issue 4. Article 101385. DOI: 10.1016/j.giq.2019.06.002.
14. *Milmo D.* Elon Musk says all human data for AI training “exhausted” // *The Guardian.* 2025. Jan 9. URL: <https://www.theguardian.com/technology/2025/jan/09/elon-musk-data-ai-training-artificial-intelligence>.

A digital lake can be filled with synthetic data when the information it contains is generated by artificial intelligence that creates new data or transforms existing information. It is essential to implement methods and technical tools to distinguish between real and synthetic information within the digital lake.

A digital lake may contain information of a secondary nature. For example, it might include an image of a bus (source information) with people inside or a person crossing the road in front of the bus (secondary information). Consequently, artificial intelligence gains access not only to permitted information but also to related information, which is not entirely correct.

Finally, a digital lake can be entirely constructed from synthetic data.

Thus, we can draw several conclusions regarding the potential use of the data lakes being developed in Russia for generation of synthetic data for the purposes of public administration and legal regulation in this field.

1. The use of synthetic data presents significant opportunities for the exploration, development, and assessment of artificial intelligence technologies in the realm of public administration. Considering the prevailing trend of enhancing state capabilities with generative AI technologies to improve effectiveness, synthetic data derived from government sources can serve as a relatively secure foundation for the advancement of technological solutions and scientific research in this field.

2. The legal framework governing synthetic data necessitates additional research and the development of legal structures. At a minimum, the following is required:

- differentiation and correlation of the methods of pseudonymization, anonymization, and depersonalization (guaranteed depersonalization) of data, along with algorithms and tools for creating synthetic data;
- the determination of the parameters for the protective regime of synthetic data is essential, particularly when its use may pose risks of harm to individuals or the state, especially in the case of or semisynthetic data;
- the delineation of access to national data lakes for the creation of synthetic data, along with guidelines for the subsequent reuse of that synthetic data.

3. An important aspect of introducing synthetic data into the public administration system is the establishment of procedures and guidelines for evaluating algorithms used to generate synthetic data. This evaluation should focus on the accuracy, relevance to the intended purposes, and elimination of potential errors and inconsistencies.

#### References

1. Filyushchenko L. I. *Tsifrovoy profil' rabotnika: problemy zashchity personal'nykh dannykh* [Digital Employee Profiling: Problems of Personal Data Protection]. *Zakony Rossii: opyt, analiz, praktika* – Laws of Russia: Experience, Analysis, Practice. 2024. Issue 11. Pp. 93–97. (In Russ.).
2. Kinyakina E. *Obezlichennye dannye priravnyayut k personal'nyim* [Depersonalized Data to Be Equated to Personal Data]. *Vedomosti.* 2021, March 12. Available at: <https://www.vedomosti.ru/technology/articles/2021/03/11/861164-obezhichennye-dannie>. (In Russ.).
3. Sazonova M. *Obezlichivanie dannykh: sokhranenie balansa mezhdru pravami grazhdan i razvitiem innovatsiy* [Depersonalization of Data: Maintaining a Balance between the Rights of Citizens and the Development of Innovation]. *Garant.* 2021, May 24. Available at: <https://www.garant.ru/news/1464529/>. (In Russ.).
4. Solovkin S. *Publichno-pravovoe regulirovanie protsessov avtomatizirovannogo sbora dannykh o cheloveke v kontekste tsifrovoy transformatsii: diss. ... kand. yurid. nauk* [Public Law Regulation of Automated Collection of Data on a Person in the Context of Digital Transformation: Cand. jurid. sci. diss.]. Moscow, 2024. 278 p. (In Russ.).
5. Abramov M. Synthetic Data Definition: Pros and Cons. *KeyMakr.* 2024. Available at: <https://keymakr.com/blog/synthetic-data-definition-pros-and-cons/>. (In Eng.).
6. Bullman H. The DoD and U.S. Census Bureau Turn to Data Lakes for Modernization and Greater Insights. *The Government Technology Insider.* 2020. Available at: <https://governmenttechnologyinsider.com/the-dod-and-u-s-census-bureau-turn-to-data-lakes-for-modernization-and-greater-insights/>. (In Eng.).
7. Bundin M., Martynov A., Aliev Y., Kutuev E. Legal Aspects of the Use of AI in Public Sector. *Digital Transformation and Global Society.* Springer, Cham, 2018. Vol. 858. Pp.171–180. (Series: Communications in Computer and Information Science). DOI: 10.1007/978-3-030-02843-5\_14. (In Eng.)
8. Dixon J. Pentaho, Hadoop, and Data Lakes. *WordPress.com.* 2010. Available at: <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes>. (In Eng.).
9. Farrell E., Minghini M. et al. *European Data Spaces – Scientific Insights into Data Sharing and Utilisation at Scale.* Luxembourg: Publications Office of the European Union, 2023. 78 p. DOI: 10.2760/400188. (In Eng.).

15. Pokotylo P. Ethical and Legal Considerations of Synthetic Data Usage. URL: <https://keymakr.com/blog/ethical-and-legal-considerations-of-synthetic-data-usage/>.
16. Stone A. Synthetic Data Supports State and Local Government AI Initiatives // StateTech. 2024. Jul 18. URL: <https://statetechmagazine.com/article/2024/07/synthetic-data-supports-ai-initiatives-for-municipalities-perfcon>.
17. Villalobos P. et al. Will we run out of data? Limits of LLM scaling based on human-generated data // ArXiv. Available at: <https://arxiv.org/abs/2211.04325>. DOI: 10.48550/arXiv.2211.04325
18. Weerts S. Generative AI in public administration in light of the regulatory awakening in the US and EU // Cambridge Forum on AI: Law and Governance. 2025. Vol. 1. Pp. 1–19. DOI: 10.1017/cfl.2024.10.

#### References

1. Filyushchenko L. I. *Tsifrovoy profil' rabotnika: problemy zashchity personal'nykh dannykh* [Digital Employee Profiling: Problems of Personal Data Protection]. *Zakony Rossii: opyt, analiz, praktika* – Laws of Russia: Experience, Analysis, Practice. 2024. Issue 11. Pp. 93–97. (In Russ.).
2. Kinyakina E. *Obezlichennye dannye priravnyayut k personal'nyim* [Depersonalized Data to Be Equated to Personal Data]. *Vedomosti*. 2021, March 12. Available at: <https://www.vedomosti.ru/technology/articles/2021/03/11/861164-obezlichennye-dannye>. (In Russ.).
3. Sazonova M. *Obezlichivanie dannykh: sokhraneniye balansa mezhdru pravami grazhdan i razvitiem innovatsiy* [Depersonalization of Data: Maintaining a Balance between the Rights of Citizens and the Development of Innovation]. *Garant*. 2021, May 24. Available at: <https://www.garant.ru/news/1464529/>. (In Russ.).
4. Solovkin S. *Publichno-pravovoe regulirovaniye protsessov avtomatizirovannogo sbora dannykh o cheloveke v kontekste tsifrovoy transformatsii: diss. ... kand. yurid. nauk* [Public Law Regulation of Automated Collection of Data on a Person in the Context of Digital Transformation: Cand. jurid. sci. diss.]. Moscow, 2024. 278 p. (In Russ.).
5. Abramov M. Synthetic Data Definition: Pros and Cons. *KeyMakr*. 2024. Available at: <https://keymakr.com/blog/synthetic-data-definition-pros-and-cons/>. (In Eng.).
6. Bullman H. The DoD and U.S. Census Bureau Turn to Data Lakes for Modernization and Greater Insights. *The Government Technology Insider*. 2020. Available at: <https://governmenttechnologyinsider.com/the-dod-and-u-s-census-bureau-turn-to-data-lakes-for-modernization-and-greater-insights/>. (In Eng.).

7. Bundin M., Martynov A., Aliev Y., Kutuev E. Legal Aspects of the Use of AI in Public Sector. *Digital Transformation and Global Society*. Springer, Cham, 2018. Vol. 858. Pp.171–180. (Series: Communications in Computer and Information Science). DOI: 10.1007/978-3-030-02843-5\_14. (In Eng.)
8. Dixon J. Pentaho, Hadoop, and Data Lakes. *WordPress.com*. 2010. Available at: <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes>. (In Eng.).
9. Farrell E., Minghini M. et al. *European Data Spaces – Scientific Insights into Data Sharing and Utilisation at Scale*. Luxembourg: Publications Office of the European Union, 2023. 78 p. DOI: 10.2760/400188. (In Eng.).
10. Inmon W. H. *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. New Jersey: Technics Publications, 2016. 166 p. (In Eng.).
11. Jälmy M., Elvander F., Waterschoot T. Multi-Channel Low-Rank Convolution of Jointly Compressed Room Impulse Responses. *IEEE Open Journal of Signal Processing*. 2024. Vol. 5. Pp. 850–857. DOI: 10.1109/OJSP.2024.3410089. (In Eng.).
12. Kitchin R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications Ltd., 2014. 240 p. DOI: 10.4135/9781473909472. (In Eng.).
13. Mergel I., Edelmann N., Haug N. Defining Digital Transformation: Results from Expert Interviews. *Government Information Quarterly*. 2019. Vol. 36. Issue 4. Article 101385. DOI: 10.1016/j.giq.2019.06.002. (In Eng.).
14. Milmo D. Elon Musk Says All Human Data for AI Training 'Exhausted'. *The Guardian*. 2025. Available at: <https://www.theguardian.com/technology/2025/jan/09/elon-musk-data-ai-training-artificial-intelligence>. (In Eng.).
15. Pokotylo P. Ethical and Legal Considerations of Synthetic Data Usage. *KeyMakr*. 2024. Available at: <https://keymakr.com/blog/ethical-and-legal-considerations-of-synthetic-data-usage>. (In Eng.).
16. Stone A. Synthetic Data Supports State and Local Government AI Initiatives. *The Statetechmagazine*. 2024. Available at: <https://statetechmagazine.com/article/2024/07/synthetic-data-supports-ai-initiatives-for-municipalities-perfcon>. (In Eng.).
17. Villalobos P., Ho A., Sevilla J., Besiroglu T., Heim L., Hobbhahn M. Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data. 2022. *ArXiv*. Available at: <https://arxiv.org/abs/2211.04325>. DOI: 10.48550/arXiv.2211.04325. (In Eng.).
18. Weerts S. Generative AI in Public Administration in Light of the Regulatory Awakening in the US and EU. *Cambridge Forum on AI: Law and Governance*. 2025. Vol. 1. Pp. 1–19. DOI: 10.1017/cfl.2024.10. (In Eng.).

10. Inmon W. H. *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. New Jersey: Technics Publications, 2016. 166 p. (In Eng.).
11. Jälmy M., Elvander F., Waterschoot T. Multi-Channel Low-Rank Convolution of Jointly Compressed Room Impulse Responses. *IEEE Open Journal of Signal Processing*. 2024. Vol. 5. Pp. 850–857. DOI: 10.1109/OJSP.2024.3410089. (In Eng.).
12. Kitchin R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications Ltd., 2014. 240 p. DOI: 10.4135/9781473909472. (In Eng.).
13. Mergel I., Edelmann N., Haug N. Defining Digital Transformation: Results from Expert Interviews. *Government Information Quarterly*. 2019. Vol. 36. Issue 4. Article 101385. DOI: 10.1016/j.giq.2019.06.002. (In Eng.).
14. Milmo D. Elon Musk Says All Human Data for AI Training 'Exhausted'. *The Guardian*. 2025. Available at: <https://www.theguardian.com/technology/2025/jan/09/elon-musk-data-ai-training-artificial-intelligence>. (In Eng.).
15. Pokotylo P. Ethical and Legal Considerations of Synthetic Data Usage. *KeyMakr*. 2024. Available at: <https://keymakr.com/blog/ethical-and-legal-considerations-of-synthetic-data-usage>. (In Eng.).
16. Stone A. Synthetic Data Supports State and Local Government AI Initiatives. *The Statetechmagazine*. 2024. Available at: <https://statetechmagazine.com/article/2024/07/synthetic-data-supports-ai-initiatives-for-municipalities-perfcon>. (In Eng.).
17. Villalobos P., Ho A., Sevilla J., Besiroglu T., Heim L., Hobbhahn M. Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data. 2022. *ArXiv*. Available at: <https://arxiv.org/abs/2211.04325>. DOI: 10.48550/arXiv.2211.04325. (In Eng.).
18. Weerts S. Generative AI in Public Administration in Light of the Regulatory Awakening in the US and EU. *Cambridge Forum on AI: Law and Governance*. 2025. Vol. 1. Pp. 1–19. DOI: 10.1017/cfl.2024.10. (In Eng.).

#### References in Russia

1. Филлющенко Л. И. Цифровой профиль работника: проблемы защиты персональных данных // Законы России: опыт, анализ, практика. 2024. № 11. С. 93–97.
2. Кинякина Е. Обезличенные данные приравняют к персональным // Вестник. 2021. 12 марта. URL: <https://www.vedomosti.ru/technology/articles/2021/03/11/861164-obezlichennye-dannye>.
3. Сазонова М. Обезличивание данных: сохранение баланса между правами граждан и развитием инноваций // Гарант.ру. 2021. 24 мая. URL: <https://www.garant.ru/news/1464529/>.
4. Соловкин С. В. Публично-правовое регулирование процессов автоматизированного сбора данных о человеке в контексте цифровой трансформации: дис. ... канд. юрид. наук. М., 2025. 278 с.

5. Abramov M. Synthetic Data Definition: Pros and Cons. URL: <https://keymakr.com/blog/synthetic-data-definition-pros-and-cons/>.
6. Bullman H. The DoD and U.S. Census Bureau Turn to Data Lakes for Modernization and Greater Insights // Government Technology Insider. 2020. Aug. 6. URL: <https://governmenttechnologyinsider.com/the-dod-and-u-s-census-bureau-turn-to-data-lakes-for-modernization-and-greater-insights/>.
7. Bundin M., Martynov A., Aliev Y., Kutuev E. Legal Aspects of the Use of AI of Public Sector // Digital Transformation and Global Society (Series: Communications in Computer and Information Science). Vol. 858. Springer Cham, 2018. Pp. 171–180. DOI: 10.1007/978-3-030-02843-5\_14.
8. Dixon J. Pentaho, Hadoop, and Data Lakes. URL: <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes>.
9. Farrell E. et al. *European Data Spaces – Scientific Insights into Data Sharing and Utilisation at Scale*. Luxembourg: Publications Office of the European Union, 2023. 78 p. DOI: 10.2760/400188.
10. Inmon W. H. *Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump*. N.J.: Technics Publications, 2016.166 p.
11. Jälmy M., Elvander F., Waterschoot T. Multi-Channel Low-Rank Convolution of Jointly Compressed Room Impulse Responses // IEEE Open Journal of Signal Processing. 2024. Vol. 5. Pp. 850–857. DOI: 10.1109/OJSP.2024.3410089.
12. Kitchin R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE Publications Ltd., 2014. 240 p. DOI: 10.4135/9781473909472.
13. Mergel I., Edelmann N., Haug N. Defining Digital Transformation: Results from Expert Interviews // Government Information Quarterly. 2019. Vol. 36. Issue 4. Article 101385. DOI: 10.1016/j.giq.2019.06.002.
14. Milmo D. Elon Musk says all human data for AI training "exhausted" // The Gardian. 2025. Jan 9. URL: <https://www.theguardian.com/technology/2025/jan/09/elon-musk-data-ai-training-artificial-intelligence>.
15. Pokotylo P. Ethical and Legal Considerations of Synthetic Data Usage. URL: <https://keymakr.com/blog/ethical-and-legal-considerations-of-synthetic-data-usage/>.
16. Stone A. Synthetic Data Supports State and Local Government AI Initiatives // StateTech. 2024. Jul 18. URL: <https://statetechmagazine.com/article/2024/07/synthetic-data-supports-ai-initiatives-for-municipalities-perfcon>.
17. Villalobos P. et al. Will we run out of data? Limits of LLM scaling based on human-generated data // ArXiv. Available at: <https://arxiv.org/abs/2211.04325>. DOI: 10.48550/arXiv.2211.04325
18. Weerts S. Generative AI in public administration in light of the regulatory awakening in the US and EU // Cambridge Forum on AI: Law and Governance. 2025. Vol. 1. Pp. 1–19. DOI: 10.1017/cfl.2024.10.

**Информация об авторах:**

**А. В. Мартынов**

Доктор юридических наук, профессор,  
главный научный сотрудник, заведующий кафедрой  
административного и финансового права  
Национальный исследовательский Нижегородский  
государственный университет им. Н. И. Лобачевского  
603022, Россия, г. Нижний Новгород, пр-т Гагарина, 23

ORCID: 0000-0002-6322-818X  
ResearcherID: U-7632-2017

Статьи в БД Scopus / Web of Science:

DOI: 10.1007/978-3-319-62530-0\_11  
DOI: 10.1145/3463677.3463758  
DOI: 10.1007/978-3-030-39296-3\_21  
DOI: 10.1145/3428502.3428531  
DOI: 10.1145/3396956.3397866  
DOI: 10.1007/978-3-030-13283-5\_15  
DOI: 10.1109/ICEDEG.2019.8734367  
DOI: 10.1007/978-3-030-02843-5\_14  
DOI: 10.1145/3047273.3047337  
DOI: 10.1007/978-3-319-69784-0\_20  
DOI: 10.1007/978-3-319-49700-6\_17  
DOI: 10.1145/2912160.2912211  
DOI: 10.1145/2910019.2910068  
DOI: 10.1145/2757401.2757454  
DOI: 10.1145/2691195.2691222  
DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**Е. В. Ширеева**

Кандидат юридических наук, доцент,  
старший научный сотрудник, доцент кафедры  
административного и финансового права  
Национальный исследовательский Нижегородский  
государственный университет им. Н. И. Лобачевского  
603022, Россия, г. Нижний Новгород, пр-т Гагарина, 23

ORCID: 0000-0002-0373-438X  
ResearcherID: U-8242-2017

Статьи в БД Scopus / Web of Science:

DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**About the authors:**

**A. V. Martynov**

Lobachevsky State University of Nizhny Novgorod  
23, prospekt Gagarina, Nizhny Novgorod, 603022, Russia

ORCID: 0000-0002-6322-818X  
ResearcherID: U-7632-2017

Articles in Scopus / Web of Science:

DOI: 10.1007/978-3-319-62530-0\_11  
DOI: 10.1145/3463677.3463758  
DOI: 10.1007/978-3-030-39296-3\_21  
DOI: 10.1145/3428502.3428531  
DOI: 10.1145/3396956.3397866  
DOI: 10.1007/978-3-030-13283-5\_15  
DOI: 10.1109/ICEDEG.2019.8734367  
DOI: 10.1007/978-3-030-02843-5\_14  
DOI: 10.1145/3047273.3047337  
DOI: 10.1007/978-3-319-69784-0\_20  
DOI: 10.1007/978-3-319-49700-6\_17  
DOI: 10.1145/2912160.2912211  
DOI: 10.1145/2910019.2910068  
DOI: 10.1145/2757401.2757454  
DOI: 10.1145/2691195.2691222  
DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**E. V. Shireeva**

Lobachevsky State University of Nizhny Novgorod  
23, prospekt Gagarina, Nizhny Novgorod, 603022, Russia

ORCID: 0000-0002-0373-438X  
ResearcherID: U-8242-2017

Articles in Scopus/Web of Science:

DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**About the authors:**

**A. V. Martynov**

Lobachevsky State University of Nizhny Novgorod  
23, prospekt Gagarina, Nizhny Novgorod, 603022, Russia

ORCID: 0000-0002-6322-818X  
ResearcherID: U-7632-2017

Articles in Scopus / Web of Science:

DOI: 10.1007/978-3-319-62530-0\_11  
DOI: 10.1145/3463677.3463758  
DOI: 10.1007/978-3-030-39296-3\_21  
DOI: 10.1145/3428502.3428531  
DOI: 10.1145/3396956.3397866  
DOI: 10.1007/978-3-030-13283-5\_15  
DOI: 10.1109/ICEDEG.2019.8734367  
DOI: 10.1007/978-3-030-02843-5\_14  
DOI: 10.1145/3047273.3047337  
DOI: 10.1007/978-3-319-69784-0\_20  
DOI: 10.1007/978-3-319-49700-6\_17  
DOI: 10.1145/2912160.2912211  
DOI: 10.1145/2910019.2910068  
DOI: 10.1145/2757401.2757454  
DOI: 10.1145/2691195.2691222  
DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
**DOI: 10.1109/ICEDEG58167.2023.10121979**

**E. V. Shireeva**

Lobachevsky State University of Nizhny Novgorod  
23, prospekt Gagarina, Nizhny Novgorod, 603022, Russia

ORCID: 0000-0002-0373-438X  
ResearcherID: U-8242-2017

Articles in Scopus/Web of Science:

DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**Информация об авторах:**

**А. В. Мартынов**

Доктор юридических наук, профессор,  
главный научный сотрудник, заведующий кафедрой  
административного и финансового права  
Национальный исследовательский Нижегородский  
государственный университет им. Н. И. Лобачевского  
603022, Россия, г. Нижний Новгород, пр-т Гагарина, 23

ORCID: 0000-0002-6322-818X  
ResearcherID: U-7632-2017

Статьи в БД Scopus / Web of Science:

DOI: 10.1007/978-3-319-62530-0\_11  
DOI: 10.1145/3463677.3463758  
DOI: 10.1007/978-3-030-39296-3\_21  
DOI: 10.1145/3428502.3428531  
DOI: 10.1145/3396956.3397866  
DOI: 10.1007/978-3-030-13283-5\_15  
DOI: 10.1109/ICEDEG.2019.8734367  
DOI: 10.1007/978-3-030-02843-5\_14  
DOI: 10.1145/3047273.3047337  
DOI: 10.1007/978-3-319-69784-0\_20  
DOI: 10.1007/978-3-319-49700-6\_17  
DOI: 10.1145/2912160.2912211  
DOI: 10.1145/2910019.2910068  
DOI: 10.1145/2757401.2757454  
DOI: 10.1145/2691195.2691222  
DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979

**Е. В. Ширеева**

Кандидат юридических наук, доцент,  
старший научный сотрудник, доцент кафедры  
административного и финансового права  
Национальный исследовательский Нижегородский  
государственный университет им. Н. И. Лобачевского  
603022, Россия, г. Нижний Новгород, пр-т Гагарина, 23

ORCID: 0000-0002-0373-438X  
ResearcherID: U-8242-2017

Статьи в БД Scopus / Web of Science:

DOI: 10.1007/978-3-030-67238-6\_7  
DOI: 10.1145/3209415.3209448  
DOI: 10.1007/978-3-319-69784-0\_12  
DOI: 10.1007/978-3-031-04238-6\_7  
DOI: 10.1109/ICEDEG58167.2023.10121979