



**Алексей Игоревич
ХИМЧЕНКО,**

доцент кафедры
информационного права
и цифровых технологий
Университета имени
О.Е. Кутафина (МГЮА),
кандидат юридических наук
aihimchenko@msal.ru
125993, Россия, г. Москва,
ул. Садовая-Кудринская, д. 9

К вопросу о роли данных в обеспечении доверенного функционирования систем искусственного интеллекта

Аннотация. Статья посвящена исследованию проблематики доверенного функционирования систем искусственного интеллекта. При проведении исследования использовались различные методы, в том числе общенаучный, системный метод, методы обобщения и правового анализа. Активное развитие решений в сфере искусственного интеллекта оказывает значительное влияние на структуру общественных отношений, происходит фундаментальная перестройка механики функционирования многих общественных и деловых процессов. В указанных условиях, с учетом потенциальных рисков использования, актуализируется вопрос доверенного функционирования систем искусственного интеллекта как важнейшего условия, определяющего возможность применения этих систем при решении задач обработки данных.

Исследуются риски и вопросы безопасности функционирования систем ИИ. Анализируются ключевые структурные элементы, образующие систему доверия к системам ИИ, способы обеспечения доверия, характеристики систем ИИ. Анализируется роль данных в механике функционирования систем ИИ. Выделяются и структурируются направления внедрения доверенных технологий ИИ, рассматриваются вопросы поиска практических решений по повышению безопасного и доверенного функционирования систем ИИ.

Ключевые слова: искусственный интеллект, системы искусственного интеллекта, доверие, цифровая среда, стандарт, безопасность, надежность, данные, оборот данных, жизненный цикл, доверенные технологии, нейросети

DOI: 10.17803/2311-5998.2026.142.6.086-094

Aleksey I. KHMCHENKO,

Associate Professor of the Department of Information Law and Digital Technologies
of the Kutafin Moscow State Law University (MSAL), Cand. Sci. (Law)
aihimchenko@msal.ru

9, ul. Sadovaya-Kudrinskaya, Moscow, Russia, 125993

On the Issue of Trustworthy Functioning of Artificial Intelligence Systems and the Role of Data

Abstract. The article is devoted to the study of the problems of trusted functioning of artificial intelligence systems. Methods: analysis, system method, idealization, legal modeling, hypothesis building, generalization and system

analysis. Results. The active development of artificial intelligence is redirecting key technological approaches from digital transformation to AI transformation.

In these conditions, taking into account the potential risks of use, the issue of trusted operation of artificial intelligence systems becomes relevant, as it is a crucial condition that determines the possibility of using these systems for data processing tasks.

The article explores the risks and security issues associated with the operation of AI systems. It analyzes the key structural elements that form the system of trust in AI systems, the methods of ensuring trust, and the characteristics of AI systems. The article also examines the role of data in the operation of AI systems. The article highlights and structures the areas of implementation of trusted AI technologies, and discusses the search for practical solutions to improve the secure and trusted operation of AI systems.

Keywords: artificial intelligence, artificial intelligence systems, trust, digital environment, standard, security, reliability, data, lifecycle, trusted technologies, neural networks

Введение

На современном этапе технологического развития существенное внимание отводится тематике искусственного интеллекта (ИИ), что преломляет отраслевой фокус внимания от цифровой трансформации к ИИ-трансформации. Использование ИИ для поиска и генерации информации, обработки данных и аналитики значительно влияет на эффективность и трудозатратность большинства существующих процессов — расширяет доступность цифровых услуг, сокращает большинство географических ограничений, способствует решению экологических, социальных и управленческих задач.

Внедрение ИИ в экономику, социальную сферу и государственное управление уже сегодня приносит определенный эффект¹ и, по экспертным оценкам, может достичь показателя в 4 % ВВП к 2028 г.² При этом ожидаемый экономический эффект к 2030 г. оценивается в 7,9—12,8 трлн руб. в год³. Например, только в здравоохранении прогнозируется эффект в 65 млрд руб. к 2030 г. посредством развития различных групп продуктов, таких как цифровые ассистенты, клиническая сводка, помощник врача, экспертный контроль⁴.

Помимо прямых экономических эффектов, к позитивным факторам относят также повышение качества обработки и аналитики информации, рост динамики и эффективности технологических процессов, расширение географии оказания

¹ URL: <http://government.ru/news/55306/> (дата обращения: 01.03.2026).

² URL: https://yakov.partners/upload/iblock/c5e/c8t1wrkdne5y9a4nqlcderalwny7xh4/20231218_AI_future.pdf?ysclid=lvj59xugqn182161365 (дата обращения: 01.03.2026).

³ URL: <https://yakovpartners.ru/publications/ai-2025/> (дата обращения: 01.03.2026).

⁴ URL: <https://yakovpartners.ru/publications/ai-healthcare/> (дата обращения: 01.03.2026).



услуг, улучшение ключевых метрик в процессах поддержки принятия решений, персонализацию и адресность, повышение безопасности и качества жизни, снижение издержек в ключевых отраслях экономики.

В то же время, формируя значительные предпосылки для фундаментальных преобразований в различных сферах общественных отношений, ИИ-решения вызывают и определенные опасения в части влияния на общественные процессы, культуру⁵, возможного нанесения вреда при обучении⁶, качества образования⁷.

Рассматривая вопросы изменения роли человека и права в информационных правоотношениях, В. Б. Наумов отмечает, что «несмотря на мизерный в исторических масштабах период проникновения ИИ в нашу жизнь, некоторые проводимые исследования дают основания для подтверждения гипотез обширного негативного воздействия ИИ», поднимая при этом проблематику критического мышления человеческого общества, изменения соответствующих навыков и воли⁸.

Отчасти подтверждая наметившиеся изменения в конфигурации человеческих навыков, исследование РОМИР фиксирует нисходящую тенденцию использования офисных программ (Word, Excel, PowerPoint и аналогов) и языков программирования (Python, SQL, C++ и др.) при увеличении динамики использования ИИ и нейросетей⁹.

Вместе с тем в отраслевых исследованиях¹⁰ экспертами также выявляются и другие риски применения ИИ-решений:

- рост количества некачественного контента в Интернете;
- принятие решений на основе недостоверных ответов сервиса;
- видоизменение рынка труда;
- стимулирование цифрового мошенничества;
- нарушение принятых в обществе норм.

Рассматривая роль и влияние ИИ на общественные отношения, следует отметить, что важным элементом цивилизованного развития технологий (и одновременно сдерживающим потенциальные негативные эффекты развития) традиционно выступает правовое регулирование. Однако и в этой плоскости отмечаются скудная и требующая ревизии система правового регулирования¹¹, необходимость поиска правовых конструкций регулирования¹², а вопросы, связанные с

⁵ Василевская Л. Ю. Кодекс этики для искусственного интеллекта: юридический миф и реальность // Гражданское право. 2023. № 2. С. 20.

⁶ Архипов В. В., Наумов В. Б., Старостенко В. А., Нестеров А. А. Использование искусственного интеллекта в высшем образовании: опыт СПбГУ // Труды по интеллектуальной собственности (Works on Intellectual Property). 2025. Т. 52. № 1. С. 128.

⁷ URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/iskusstvennyi-intellekt-v-shkole-dobro-ili-zlo> (дата обращения: 01.03.2026).

⁸ Наумов В. Б. Право и «Батлерианский джихад»: размышления об угрозах использования искусственного интеллекта // Информационное общество. 2025. № 2. С. 140.

⁹ Отчет «Информационные технологии в восприятии россиян — 2025». М., 2026. С. 64.

¹⁰ URL: <https://yakovpartners.ru/publications/gen-ai-regulation/> (дата обращения: 01.03.2026).

¹¹ Наумов В. Б. Указ. соч. С. 136.

¹² Батурич Ю. М., Полубинская С. В. Искусственный интеллект: правовой статус или правовой режим? // Государство и право. 2022. № 10. С. 152.

правовым статусом ИИ, рассматриваются в контексте защиты человека от неконтролируемого влияния ИИ и неизменности гарантий конституционных прав и свобод человека¹³.

В этой связи нельзя не отметить создание Центра развития искусственного интеллекта¹⁴, в задачи которого входят «нормативно-правовое регулирование обеспечения безопасности технологий ИИ, а также тиражирование лучших практик, к которым отнесены решения, доказавшие свою эффективность и безопасность». Очевидно, что выделение для тиражирования безопасных практик подчеркивает значимую роль аспекта безопасности в функционировании систем ИИ.

Значимость проблематики безопасности обусловлена в том числе большим набором возможностей для атак на системы искусственного интеллекта, среди которых отравление данных, атака инверсии модели, атака градиентного спуска и подмена модели, подмена данных и др.¹⁵ Не углубляясь в специфику защиты систем ИИ, стоит отметить, что столь широкий спектр потенциальных атак предопределяет необходимость классификации атак на системы ИИ.

Таким образом, лишь поверхностный анализ специфики функционирования ИИ-систем в контексте их влияния на общественные отношения позволяет отметить наличие комплексного рискового потенциала, что, в свою очередь, говорит о необходимости структурирования и систематизации соответствующих рисков.

Вместе с тем, помимо рисков результирующего воздействия ИИ-решений на общественные отношения, стоит выделить также значимую технологическую роль данных в механике функционирования систем ИИ и специфику их использования.

Поскольку технологически модели машинного обучения выполняют аппроксимацию требуемой функции по таблице ее значений, построенной на основе заданного обучающего набора наблюдений, очевидно, что именно данные, их наборы и структура являются ключевым элементом ИИ-систем и залогом их качественного и продуктивного функционирования.

И здесь важную роль играют такие свойства данных, как качество, адекватность данных запросу (алгоритм может быть обучен на данных, ориентированных на другие задачи) и их достоверность (не зря в рамках работы над национальным проектом «Экономика данных» предусматривалась проработка мероприятий по выявлению недостоверных данных, в том числе генерируемых с использованием ИИ¹⁶).

Другой важной составляющей механики использования данных является правомерность их использования, поскольку определенные категории данных, необходимые для обучения нейронных сетей, не могут передаваться сторонним сервисам в силу обязательств по конфиденциальности (например коммерческой, адвокатской тайне и др.).

¹³ Карцхия А. А. Правовая охрана достижений искусственного интеллекта // ИС. Авторское право и смежные права. 2024. № 4. С. 4.

¹⁴ Дмитрий Григоренко: Правительство создает Центр развития искусственного интеллекта // URL: <http://government.ru/news/55306/> (дата обращения: 01.03.2026).

¹⁵ (Не)доверенный ИИ: обучать нельзя запретить // URL: <https://kryptonite.ru> <https://kryptonite.ru/articles/expert-discussion-of-trusted-ai/> (дата обращения: 01.03.2026).

¹⁶ URL: <http://government.ru/news/50469/> (дата обращения: 01.03.2026).



Нередко возникает и такая уникальная особенность технологий ИИ, как «проблема соблюдения баланса между требованиями по защите персональных данных и необходимостью их использования для обучения систем ИИ»¹⁷.

Доверенное функционирование

Любое взаимодействие, в том числе в коммуникационной и цифровой среде, в целом возможно лишь при выполнении определенного набора фундаментальных принципов, одним из которых является доверие. Не является исключением и отрасль ИИ, конструктивное развитие и совершенствование которой в условиях все более глубокой интеграции алгоритмов в общественные отношения, равно как и любых других прорывных технологий, возможно лишь в состоянии доверия, при доверенном «сотрудничестве» с пользователями и потребителями.

Вместе с тем стремительность и интенсивность процессов цифровой трансформации оказывают значительное влияние на степень доверия, «подрывая доверие людей к цифровым технологиям и к информационным коммуникациям»¹⁸. Более того, ситуация развивается в период, когда одновременно фиксируется тенденция снижения доверия и к праву, как к основному социальному регулятору¹⁹.

В указанных условиях формирования тенденции к снижению доверия ключевые документы по тематике ИИ предполагают доверенный вектор развития соответствующих технологий. Так, Кодекс этики в сфере искусственного интеллекта²⁰ рекомендует акторам ИИ проводить мероприятия, направленные на повышение уровня доверия и осведомленности пользователей и общества в целом.

В то же время Стратегия развития искусственного интеллекта²¹ (далее — Стратегия) фиксирует, что отсутствие понимания того, как искусственный интеллект достигает результатов, является одной из причин низкого уровня доверия к соответствующим технологиям и может стать препятствием для их развития, а к основным задачам развития отнесено обязательное внедрение доверенных технологий искусственного интеллекта в областях его использования, где может быть нанесен ущерб безопасности Российской Федерации.

При этом к показателям, характеризующим достижение целей Стратегии, отнесено достижение роста показателя уровня доверия граждан к соответствующим технологиям в 2030 г. до 80 %.

Не ставя целью на данном этапе исследование гносеологической стороны понятия доверия и его сущности в целом, попробуем понять, что лежит в основе

¹⁷ Наумов В. Б., Тютюк Е. В. Правовые проблемы машинного обучения // Образование и право. 2020. № 6. С. 220.

¹⁸ Наумов В. Б. Указ. соч. С. 135.

¹⁹ Polyakova T. A., Naumov V. B., Minbaleev A. V. (2022). Trust in the law during the digital transformation // Государство и право. 2022. № 11. С. 140.

²⁰ URL: <https://a-ai.ru> по состоянию на 10.01.2022 (дата обращения: 01.03.2026).

²¹ Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» // СЗ РФ. 2019. № 41. Ст. 5700.

категории доверия применительно к системам ИИ, каковы подходы к обеспечению доверенного функционирования систем ИИ и роли данных при этом.

Несмотря на всю ширину спектра общественных отношений, к которым применяется категория доверия, и частоту такого применения, смысл, как правило, определяется текущим контекстом. При этом в условиях критического влияния на правоотношения единого правового подхода в действующем законодательстве нет²².

В технологическом контексте доверие в исследованиях рассматривается как сочетание нескольких факторов: техническое доверие к непосредственному решению (т.е. способность выполнять заявленную задачу), *доверие к данным*, лежащим в основе алгоритма, и готовность к принятию результата²³. В то же время при оценке надежности систем ИИ выделяют факторы доверия: теоретическое обоснование моделей, доказанную эффективность алгоритмов решения оптимизационных задач, *доверенный датасет*, использование доверенного ПО и аппаратных платформ²⁴.

Как видим, доверие к данным в экспертных подходах выделяется среди ключевых элементов, формирующих структуру доверия к ИИ. При этом непосредственный вопрос дает немало поводов для размышления, поскольку классическая наука обеспечивает доверие к результатам, основываясь на математических вычислениях и законах физики, в том числе времени. Большинство же алгоритмов основаны лишь на эмпирических данных и могут не иметь причинно-следственной структуры.

Основные направления внедрения доверенных технологий

Возвращаясь к вопросу обеспечения доверия к системам ИИ, обратимся к положениям Стратегии, согласно которым к основным направлениям внедрения доверенных технологий ИИ отнесены:

- внедрение доверенных технологий ИИ в программы цифровой трансформации;
- формирование компетенций, позволяющих использовать доверенные технологии ИИ и большие фундаментальные модели;
- методическое и нормативно-правовое обеспечение внедрения доверенных технологий ИИ в государственном управлении;
- определение единого критерия для координации деятельности по внедрению доверенных технологий ИИ;

²² Химченко А. И. Актуальные информационно-правовые вопросы формирования цифровой среды доверия // Правовое государство: теория и практика. 2023. № 4. С. 84.

²³ Профессор Центра ИИ «Сколтеха», заведующий лабораторией в области доверенного искусственного интеллекта Иван Тюкин рассказал «Ъ-Науке», что такое доверенный искусственный интеллект // URL: <https://www.kommersant.ru/doc/7639406> (дата обращения: 11.04.2026).

²⁴ (Не)доверенный ИИ: обучать нельзя запретить.



- формирование реестра апробированных доверенных технологий ИИ, проверенных на угрозы информационной безопасности;
- подготовка проектов по разработке и внедрению доверенных технологий ИИ в отраслях экономики и социальной сферы;
- создание инфраструктуры, необходимой для использования доверенных технологий ИИ;
- внедрение сертифицированных решений;
- обеспечение внедрения и использования доверенных технологий ИИ для выполнения текущих задач и другие мероприятия.

В основе подхода к созданию комплексной системы нормативно-правового регулирования общественных отношений, связанных с развитием и использованием технологий ИИ, определены следующие принципы:

- разработка, создание и использование технологий ИИ для тех областей их применения, в которых может быть нанесен ущерб безопасности Российской Федерации, в соответствии с требованиями информационной безопасности доверенных технологий;
- создание доверенного программного обеспечения для разработки безопасных и функционально эффективных решений в области ИИ по единым открытым стандартам.

К числу задач развития отнесено и формирование «единого пространства» доверенных технологий ИИ, продвижение отечественных доверенных технологий ИИ.

Вместе с тем в положениях Стратегии определяется, что доверенные технологии искусственного интеллекта — это технологии, отвечающие стандартам безопасности, разработанные с учетом принципов объективности, недискриминации, этичности, исключающие при их использовании возможность причинения вреда человеку и нарушения его основополагающих прав и свобод, нанесения ущерба интересам общества и государства, технологическое решение, соответствующее стандартам безопасности, разработанное с учетом объективности, непредвзятости, этичности и исключения применения для нанесения угрозы гражданам и государству. При этом выделить среди всего массива технологий ИИ доверенные не представляется возможным, так как по этим составляющим не разработаны требования и методики подтверждения требований.

Поскольку согласно Стратегии технологически ИИ состоит из информационно-коммуникационной инфраструктуры, ПО и сервисов обработки данных и принятия решений, то для обеспечения доверия ИИ к указанным сущностям необходимо сформировать и предъявить стандарты безопасности, критерии «доверенности».

Помимо этого, имеет смысл обратиться также к техническому регулированию. Так, проблематика доверия находит отражение в Национальном стандарте РФ ГОСТ Р 59276-2020 (Стандарт)²⁵, где доверие определяется как важнейшее условие при решении ответственных задач обработки данных.

²⁵ ГОСТ Р 59276-2020 «Национальный стандарт Российской Федерации. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения» (утв. и введен в действие приказом Росстандарта от 23.12.2020 № 1371-ст).

Вопросы обеспечения доверия к системам ИИ рассматриваются на основе подтверждения их качества. В соответствии со Стандартом систему подтверждения образуют существенные характеристики систем ИИ, а также факторы, приводящие к снижению качества систем ИИ. Кроме того, на обеспечение доверия оказывают влияние особенности, возникающие на различных стадиях жизненного цикла, а также уровнях архитектуры систем ИИ.

Проверка доверия к системе ИИ реализуется посредством установления подтверждения соответствия характеристик системы ИИ установленным требованиям. В связи с тем, что способность системы ИИ вызывать доверие может изменяться во времени, Стандартом рекомендуется использовать контрольные точки проверки доверия на протяжении жизненного цикла системы. В зависимости от фактора снижения качества системы ИИ Стандартом определяются способы обеспечения доверия к ней. При определении доверия систем ИИ выделяются три уровня архитектуры (физический, инфраструктурный и прикладной), для каждого из которых формируются свои существенные характеристики, к которым отнесены выполнение проверочных мероприятий, выполнение требований безопасности ИТ-инфраструктуры, подтверждения надежности и безопасности ПО.

В условиях фактического отсутствия правового регулирования ИИ следует отметить и то, что проблематика регулирования ИИ находится в фокусе внимания профильных институтов. Так, Минцифры России опубликовало для общественного обсуждения законопроект «Об основах госрегулирования сфер применения технологий искусственного интеллекта»²⁶, призванный на законодательном уровне закрепить правила разработки и использования ИИ.

Несмотря на начальную стадию проработки, нельзя не отметить внимание законодателя к проблематике обеспечения доверенного функционирования. Законопроект предусматривает разделение моделей ИИ на категории в зависимости от их происхождения и уровня доверия, при этом выделяются доверенные модели, подлежащие использованию в государственных информационных системах и на объектах критической информационной инфраструктуры.

Заключение

Тематика развития ИИ в настоящее время находится в активном фокусе внимания государства и различных институтов развития, приобретая ключевое значение в реализации и достижении стратегических целей развития, развитии отдельных секторов экономики и формировании новых технологических подходов.

Отсутствие правового регулирования на определенном этапе способствовало как динамичному развитию соответствующих технологий, так и формированию определенного рискованного потенциала от использования. Одновременно стратегический подход к развитию и применению технологий ИИ подразумевает их доверенное и безопасное использование.

Вместе с тем подходы к обеспечению доверия к системам ИИ в настоящее время находят отражение лишь на стратегическом уровне и уровне стандартизации,

²⁶ URL: <http://regulation.gov.ru/Eutities/Npa/view/166424> (дата обращения: 19.03.2026).



в то время как такие структурные составляющие, как требования, методики их подтверждения, стандарты безопасности к системам ИИ (и их составляющим), не сформированы.

Одним из ключевых (с точки зрения как технологической реализации, так и достижения указанных состояний) структурных элементов механики ИИ являются данные. И в вопросе обеспечения доверенного функционирования систем ИИ приоритетной задачей является формирование стандартов безопасности и критериев «доверенности» данных.

БИБЛИОГРАФИЯ

1. Архипов В. В., Наумов В. Б., Старостенко В. А., Нестеров А. А. Использование искусственного интеллекта в высшем образовании: опыт СПбГУ // Труды по интеллектуальной собственности (Works on Intellectual Property). — 2025. — Т. 52. — № 1.
2. Батулин Ю. М., Полубинская С. В. Искусственный интеллект: правовой статус или правовой режим? // Государство и право. — 2022. — № 10. — С. 141—154.
3. Василевская Л. Ю. Кодекс этики для искусственного интеллекта: юридический миф и реальность // Гражданское право. — 2023. — № 2. — С. 19—22.
4. Карцхия А. А. Правовая охрана достижений искусственного интеллекта // ИС. Авторское право и смежные права. — 2024. — № 4. — С. 4—16.
5. Наумов В. Б. Право и «Батлерианский джихад»: размышления об угрозах использования искусственного интеллекта // Информационное общество. — 2025. — № 2. — С. 131—144.
6. Наумов В. Б., Тютюк Е. В. Правовые проблемы машинного обучения // Образование и право. — 2020. — № 6.
7. Химченко А. И. Актуальные информационно-правовые вопросы формирования цифровой среды доверия // Правовое государство: теория и практика. — 2023. — № 4. — С. 78—86.
8. Polyakova T. A., Naumov V. B., Minbaleev A. V. Trust in the law during the digital transformation // Государство и право. — 2022. — № 11.