

СТАНДАРТИЗАЦИЯ ДАННЫХ И ПРОЦЕССОВ В ОТКРЫТОЙ ИНФОРМАЦИОННОЙ СРЕДЕ. ЧАСТЬ 2. ПЕРСПЕКТИВЫ БАНКОВСКОЙ СФЕРЫ

DATA AND PROCESSES STANDARDIZATION IN AN OPEN INFORMATION ENVIRONMENT. PART 2. FUTURE OF THE BANKING SECTOR

Бурый А.С., д-р техн. наук, ФГБУ «Институт стандартизации», Россия, г. Москва,

Лопатин И.Н., аспирант ФГБУ «Институт стандартизации», Россия, г. Москва,

Погодин И.М., аспирант ФГБУ «Институт стандартизации», Россия, г. Москва

Buryi A.S., Dr.Sc. (Technology), Russian Standardization Institute, Moscow

Lopatin I.N., graduate student, Russian Standardization Institute, Moscow

Pogodin I.M., graduate student, Russian Standardization Institute, Moscow

Статья посвящена анализу и оценке роли данных и их интеграции в распределенной информационной банковской сфере в ходе решения задач формирования баз данных, поддержки принятия решений на основе интеллектуального анализа данных.

Цель работы: оценка возможностей моделей представления данных в распределенных информационных средах и базах данных на основе интеллектуального анализа данных с привлечением нейро-сетевых инструментов для поддержки принятия решений в банковской сфере.

Методы: компаративный анализ научных публикаций в области интеграции больших языковых моделей, включая возможности интеллектуальных агентов и многоагентного представления динамики информационного взаимодействия.

Результаты: предложен концептуальный подход применения интеллектуальных агентов в ходе формирования и интеграции данных в банковской сфере для повышения обоснованности принимаемых решений, за счет повышения качества информационного взаимодействия на основе ролевого представления интеллектуальных агентов в ходе обучения доменно-специализированных языковых моделей.

Ключевые слова: информационная система, открытые системы, открытые данные, платформа, открытый банкинг, управление данными, генеративный интеллект.

Для цитирования: Бурый А.С., Лопатин И.Н., Погодин И.М. Стандартизация данных и процессов в открытой информационной среде. Часть 2. Перспективы банковской сферы // Информационно-экономические аспекты стандартизации и технического регулирования. 2026. № 1(88). С. 72–78.

The article is devoted to the analysis and assessment of the role of data and their integration in the distributed information banking sector in solving the problems of database formation, decision support based on data mining.

Purpose of the paper: to evaluate the capabilities of data representation models in distributed information environments and databases based on data mining using neuro-network tools to support decision-making in the banking sector.

Methods of study: comparative analysis of scientific publications in the field of integration of large language models, including the capabilities of intelligent agents and multi-agent representation of the dynamics of information interaction.

Study findings: a conceptual approach to the use of intelligent agents in the formation and integration of data in the banking sector is proposed to increase the validity of decisions made by improving the quality of information interaction based on the role representation of intelligent agents during the training of domain-specific language models.

Keywords: information system, open systems, open data, platform, open banking, data management, generative intelligence.

For citation: Buryi A.S., Lopatin I.N., Pogodin I.M. Data and Processes Standardization in an Open Information Environment. Part 2. Future of the Banking Sector. Information and economic aspects of standardization and technical regulation. 2026;1(88):72–78. (In Russ.)

ВВЕДЕНИЕ

В первой части данной работы [1] рассмотрена роль открытых данных в современном цифровом мире, в информационно-технологической (ИТ) среде в контексте понятий «открытые системы» – «открытые данные» – «открытая наука» [2]. Проанализированы основные тенденции и воз-

можный потенциал источников открытых данных и роль стандартизации в разработке требований к представлению данных, а также и перспектив расширения их интеграции.

Индустрия финансовых услуг переживает сейсмический сдвиг, обусловленный сближением технологических достижений, нормативных требований и меняющимися ожида-

ниями потребителей. Банк России в докладе, посвященном регулированию экосистем [3], формулирует определение трех видов экосистем: открытая, закрытая и гибридная. Главной целью экосистемы банка является омниканальность предоставляемых услуг, где создание бесшовного пути клиента поддерживается множеством каналов коммуникаций с ним [4].

В банковском секторе (БС) накоплены значительные массивы данных, отражающие бизнес-процессы, технологическую инфраструктуру, регламентную деятельность и практики реагирования на инциденты. Эти данные формируют совокупность корпоративных знаний (КЗ), на основе которых осуществляется обучение доменно-специализированных языковых моделей (Domain-Specific Language Models, DSLM). В отличие от универсальных языковых моделей, DSLM ориентированы на воспроизведение предметной логики, терминологии и типовых сценариев деятельности конкретной организации, что позволяет использовать их в прикладных задачах автоматизации и поддержки принятия решений [5].

Open Banking¹, существенно упрощенный благодаря стандартизированным интерфейсам прикладного программирования (Application Programming Interface, API), обеспечивает безопасный и бесперебойный обмен финансовыми данными и услугами (с разрешения клиентов), между традиционными финансовыми учреждениями и внешними сторонними разработчиками. В то же время API есопому стал важнейшим элементом этой модели совместной работы, предлагая стандартизированный интерфейс, который лежит в основе интеграции различных систем и обеспечивает эффективный обмен данными и функциональностью [6, 7].

Открытое банковское дело, по своей сути основанное на совместном доступе и сотрудничестве, устранило традиционные барьеры. Финансовые учреждения, которые когда-то были территориальными и защищали свои данные, теперь осознают ценность сотрудничества, используя, в частности технологии облачных вычислений, предлагая банкам облачные платформы для реализации своих моделей управления данными [8], опираясь на многообразие постоянно развивающихся облачных сервисов [9]. Сегмент BaaS (banking-as-a-service) – самый быстрорастущий в банковской сфере со среднегодовым темпом роста – 33%.²

Будущее финансовых услуг зависит от партнерских отношений, когда традиционные банковские учреждения и финтех-стартапы объединяются для создания максимальной стоимости. Этот переход к сотрудничеству – не просто

ПРИНЯТЫЕ СОКРАЩЕНИЯ

БС – банковский сектор
ЖЦД – жизненного цикла данных
ИИ – искусственный интеллект
КЗ – корпоративные знания
ПрО – предметная область
СППР – система поддержки принятия решений
API – интерфейс прикладного программирования
DSLM – доменно-специализированные языковые модели
ХaaS – X – название сервиса в облачных вычислениях

стратегический выбор; это необходимость для современной финансовой экосистемы. Глубина и широта проблем, с которыми сталкивается отрасль сегодня – технологических, нормативных и ориентированных на клиента – требуют совместного подхода. Более того, быстрые темпы технологического прогресса требуют активизации в разработке систем поддержки принятия решений (СППР) для эффективной адаптации и внедрения инноваций.

Целью настоящего исследования является оценка возможностей моделей представления данных в распределенных информационных средах и базах данных на основе интеллектуального анализа данных с привлечением нейросетевых инструментов для поддержки принятия решений в банковской сфере.

КАТЕГОРИИ И ПЛАТФОРМЫ ОТКРЫТЫХ ДАННЫХ

Корпоративные данные в Банке классифицируются по функциональным и доменным признакам и хранятся в распределенных хранилищах. Банковские транзакции, данные клиентских операций и финансовые показатели размещаются в классическом *Enterprise Data Lake*³, поддерживающем структурированные форматы хранения и аналитические витрины данных. Также данные кибербезопасности, включая журналы событий средств защиты (SIEM, SOAR, EDR и систем ИТ-мониторинга). При этом функции безопасности распределены: EDR обнаруживает угрозы на устройствах, SIEM коррелирует их, а SOAR ускоряет обработку инцидентов до минут.

Многообразие типов данных обычно включают [10]:

- структурированные;
- полуструктурированные;
- неструктурированные данные;
- метаданные;
- показатели (метрики) качества данных.

¹ Эволюция банкинга. Как и почему крупный бизнес переходит на Open API. [Электронный ресурс]. – URL: <https://sber.pro/publication/evolyutsiya-bankinga-kak-i-pochemu-krupnii-biznes-perehodit-na-open-api/> (дата обращения 27.12.2025).

² Там же.

³ Корпоративное озеро данных – это централизованное хранилище огромных объемов структурированных, полуструктурированных и неструктурированных данных.

Такие форматы данных обеспечивает возможность интегрированного использования данных как для аналитических задач, так и для обучения и эксплуатации DSLM [11], что соответствует архитектурным подходам, применяемым в Google, Microsoft и Meta при построении внутренних платформ на основе технологий искусственного интеллекта (ИИ-платформ).

В связи с появлением множества платформ для обмена данными, проанализируем наиболее популярные из них (см. табл. 1, составлено с учетом [12]).

Таблица 1

Возможности, предоставляемые платформами данных

Название платформы	Решаемые задачи
Kaggle ⁴	Предоставляет доступ к широкому спектру тематических наборов данных, предназначенных для алгоритмов искусственного интеллекта и анализа данных. Некоторые из наборов данных предназначены для сегментации изображений, обнаружения объектов и генерации изображений, а также для многих других приложений.
UCI ⁵	Платформа, которая предоставляет доступ к широкому спектру тем наборов данных для машинного обучения, включая стандартные алгоритмы классификации и кластеризации.
Облачные услуги: PaaS, IaaS	Предоставлению услуг в сфере облачных вычислений: платформа как услуга (PaaS); инфраструктура как услуга (IaaS) и другие [8, 13].
Поиск наборов данных Google (GDS)	Поисковая система, которую можно использовать для поиска наборов данных в Интернете, охватывающих широкий спектр тем, например, социальные науки или финансы.
Реестр открытых данных Amazon Web Services	Платформа, которая предоставляет доступ к широкому спектру наборов данных, размещенных Amazon, охватывающих такие темы, как климат или геопространственные данные.
Microsoft Research Open Data	Платформа, которая предоставляет доступ к широкому спектру тем наборов данных, включая, например, компьютерное зрение или обработку естественного языка.
Открытые данные Всемирного банка	Платформа, которая предоставляет доступ к наборам данных, касающихся глобального развития мира, включая, например, бедность, образование или изменение климата.

ДАННЫЕ КАК ИНФОРМАЦИОННЫЙ ПРОДУКТ

Большинство из нас уже не замечают, что выходя в открытую информационную среду, мы автоматически порождаем череду данных – то, что принято называть цифровым следом. Поставили «лайк» в интернет-магазине понравившемуся товару и Ваш голос уже участвует в информационной рекомендательной системе данного продавца. Пользуетесь банковскими интернет-услугами – формируете свой профиль, как клиент банка и т. д.

По мере того, как предприятия расширяют свои операции с данными, они сталкиваются с все более сложными задачами, связанными с интеграцией данных, управлением ими, масштабируемостью и рядом других задач, связанных с жизненным циклом данных (ЖЦД). Две архитектурные парадигмы – Data Fabric и Data Mesh – стали ведущими подходами к современному управлению данными [14].

Процесс управления данными во многом определяется их структурой. Продукты данных, Data Mesh или сеть данных представляет собой децентрализованную модель, которая согласовывает владение данными и управление ими с отдельными бизнес – областями (отчет, таблица, датасет и т. д.). Этот подход можно использовать для создания продуктов данных – надежных, заслуживающих доверия многократно используемых наборов данных.

Data Fabric (ткань данных) – это концепция распределенной работы с данными, единая и согласованная архитектура управления данными, обеспечивающая беспрепятственный доступ к данным и их обработку. Data Fabric предлагает централизованную и технологически ориентированную архитектуру, которая использует метаданные, ИИ и автоматизированные информационные системы для объединения и управления данными от разнотипных источников. Она предназначена для уменьшения разрозненности данных, путем создания виртуально-интегрированного уровня, охватывающего облачную среду, локальные сети и периферийные системы [14]. Этот подход фокусируется на согласованности, доступности данных в реально времени и централизованном управлении для обеспечения бесперебойной работы с данными.

В таблице 2 представлены определения к понятию «продукты данных», которые продолжают развиваться, в зависимости от специфики Про исследования.

В организационном контексте информационные продукты – это доступные для обнаружения, понятные, высококачественные, готовые к использованию и многократному использованию ресурсы данных, которые люди могут применять для решения различных бизнес-задач.

⁴ Kaggle. Доступно онлайн: <https://www.kaggle.com/> (дата обращения: 21.12.2025).

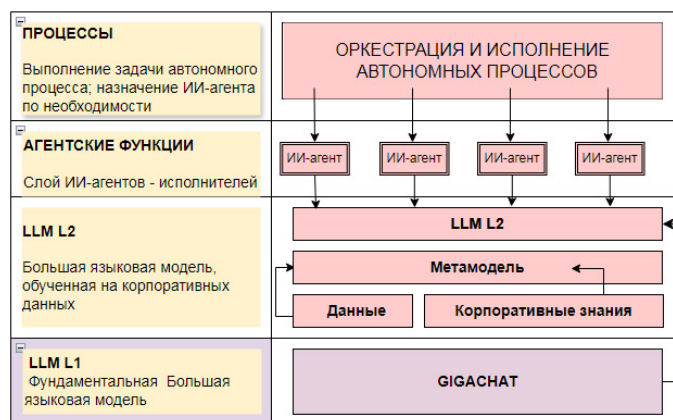
⁵ См. ГОСТ Р ИСО 9735-4-2012 Электронный обмен данными в управлении, торговле и на транспорте. М.: Стандартинформ, 2014.

Таблица 2

Сравнение подходов к понятию «данные как продукт»

Концепция	Определение
Информационный продукт	Информационный продукт – это управляемый артефакт, который удовлетворяет текущие информационные потребности и создает ценность путем преобразования и упаковки соответствующих элементов данных в удобную для использования форму.
Data Mesh	Data Mesh – это социально-техническая, децентрализованная, распределенная концепция управления корпоративными данными. Он характеризуется четырьмя принципами: данные как продукт, ориентация на Про, платформы самообслуживания и федеративное управление вычислениями.
Data Fabric	Data Fabric – это концепция дизайна для создания повторно используемых сервисов интеграции данных, конвейеров передачи данных и семантики для гибкой и интегрированной доставки данных. Он основан на анализе, создании и использовании метаданных, которые представляются в виде графа знаний.

этом различные категории данных вносят вклад в формирование языкового, логического и контекстного представления области в модели (см. рисунок). Качество этих данных является критическим фактором: при наличии неконсистентных⁸, неполных или ошибочных данных нарушается формирование устойчивых паттернов внимания и рассуждений в DSLM, что приводит к потере фокуса модели и снижению эффективности работы ИИ-агентов. Данный эффект широко описан в исследованиях Big Tech, где подчеркивается прямая зависимость качества моделей от качества обучающих данных [16].



Принципиальная схема предоставления знаний ИИ-агентам

ИСПОЛЬЗОВАНИЕ КОРПОРАТИВНЫХ ДАННЫХ ИИ-АГЕНТАМИ И ОБУЧЕНИЕ DSLM

Доступ к корпоративным данным для DSLM и ИИ-агентов реализуется через специализированную контекстную платформу (MCP-сервер)⁶, обеспечивающую управляемый и ролевой доступ к данным, а также набор утилит для их агрегации, фильтрации и предварительной обработки. DSLM в данной архитектуре используются ИИ-агентами [15] в качестве когнитивного слоя, обеспечивающего интерпретацию контекста, формирование гипотез и принятие решений, тогда как сами агенты выполняют действия, направленные на автоматизацию процессов и снижение доли рутинных операций, ранее выполняемых человеком.

Корпоративные данные применяются как при обучении DSLM, так и в процессе интерфейса⁷ через механизмы *retrieval*, т.е. когда для генерации ответов в большой языковой модели осуществляется дополнительный поиск информации во внешних источниках (файлы, базы данных определенных Про, интернет и другие источники). При

⁶ Протокол Model Context Protocol (MCP) – открытый стандарт для интеграции языковых моделей с внешними сервисами и источниками данных.

⁷ Процесс интерфейса – это совокупность средств, методов и правил, обеспечивающих взаимодействие (обмен информацией, управление, контроль) между пользователем и устройством.

Представим процессы взаимодействия агентов в виде следующей модели:

$$AM = (A, R, St_{ORG}), \tag{1}$$

где $A = \{a\}$ – множество ролей ИИ-агентов; $St_{ORG} = \{st_j\}$, $j = \overline{1, S}$ – множество информационных структур, соответствующих различным этапам ЖЦД; R – семейство базовых отношений между агентами, включающее особенности информационного взаимодействия, как между ИИ-агентами, так и внешней средой. Распределение ролей ИИ-агентов может строиться относительно категорий доменных знаний и данных для DSLM-моделей, представленных в табл. 3.

Взаимодействие (*Int*) агентов из множества A представим кортежем вида:

$$Int = \langle A, T, Sc \rangle, \tag{2}$$

где $T = \{T_i\}$ – множество типов агентов, включая агентов контроля состояния, исполнителей (работа с форматами данных), координаторов, обработки данных и другие; Sc –

⁸ Ситуация, когда одни и те же сущности отражены по-разному (на уровне данных) в подсистемах.

сценарии (программы), сетевые графики взаимодействия агентов, причем

$$Sc=(Com, \pi), \tag{3}$$

где *Com* – множество коммуникативных действий между агентами в соответствии со структурой, представленной на рис. 1; π – множество протоколов типовых действий (регистрация или измерение, запись информации, копирование данных, передача данных, отображение и др.).

Так, для ИИ-агента обеспечения связи с внешними базами данных (БД) на этапе обучения любой протокол $\pi_k \in \pi$ может иметь вид:

$$\pi_k = (\langle \text{подбор источников данных } \in \text{БД} \rangle; \langle \text{передача метаданных} \rangle; \langle \text{формирование типовых запросов} \rangle; \dots), k \in K, \tag{4}$$

где *K* – множество протоколов; БД – множество баз данных, доступных для использования в работе ИИ-агентов на любом информационном уровне.

Таким образом, любой протокол вида (4) – это набор действий, а само действие можно представить как отражение наблюдаемого агентом подмножества состояний окружающей среды (*E*) во множество возможных его ответных действий (реакций) – \bar{D} :

$$action : E \rightarrow \bar{D}. \tag{5}$$

Динамику изменения состояний окружающей среды под воздействием агента будем представлять следующей функцией:

$$\varepsilon : E^S \times \bar{D} \rightarrow 2^E. \tag{6}$$

Взаимодействие агента со средой представляет собой последовательную смену состояний, отражаемую в виде протокола (4), представим в виде последовательности смены состояний многоагентной системы в результате определенного действия агента из множества, задаваемого системой отображений вида (5). В итоге любую программу действий по преобразованию данных представим как последовательность вида «состояние-действие» для реализации заданных целей:

$$Prog : e_0 \xrightarrow{a_0} e_1 \xrightarrow{a_2} e_2 \dots \xrightarrow{a_k} e_k, \tag{7}$$

где e_i – состояния информационной системы, а $a_i, i = 0, 1, 2, \dots$ – действия ИИ-агентов, вызвавшие смену состояний, а в целом программа (7) фиксируется соответствующим протоколом вида (4).

Эффективность выполнения программ вида (7) или заданных сценариев, как совокупности программ, обычно оцениваются заданными метриками качества. Для измерительных задач это могут быть точности измерительных данных, для расчетных функций – точности преобразований (обработки данных), для СППР – уровни риска при выборе решений и другие.

Таблица 3

Категории данных и их применение при обучении DSLM

Категория данных	Характеристика	Основные форматы	Роль в обучении DSLM
Транзакционные данные	Банковские операции и финансовые события	Структурированные	Формирование понимания бизнес-логики и причинно-следственных связей
Данные клиентских взаимодействий	История обращений и коммуникаций	Полуструктурированные, текстовые	Обучение контекстному анализу и сценарному reasoning
Данные кибербезопасности	События средств защиты и мониторинга	Полуструктурированные, потоковые	Обучение интерпретации инцидентов и выявлению аномалий
Инфраструктурные данные	Метрики и log-файлы ИТ-систем	Полуструктурированные	Формирование представлений о работе инфраструктуры
Регламентные и нормативные данные	Политики, стандарты, инструкции	Неструктурированные	Обучение нормативной логике и корпоративной терминологии
Исторические данные инцидентов	Отчеты расследований	Текстовые, структурированные	Обучение модели анализу техник и тактик в инцидентах
Метаданные и каталоги	Описания источников	Структурированные	Информация по данным источникам
Данные качества данных	Метрики и результаты проверок	Структурированные	Обеспечение устойчивости рассуждений и контроля фокуса модели
Синтетические данные	Сгенерированные синтетические сценарии	Текстовые, структурированные	Дополнение обучающих выборок и моделирование редких случаев

Для управления данными в целом необходимо четко определить, какие роли ИИ-агентов важны для предоставления и обработки данных, и как эти роли должны быть включены в процесс принятия решений. При этом в СППР затрагиваются такие аспекты, как управление качеством данных, управление доступом к данным, общее управление данными и управление ЖЦД. Кроме того, ключевой задачей является управление метаданными для наборов данных о своей ПрО, что также важно в контексте суверенитета и качества данных.

ЗАКЛЮЧЕНИЕ

Ключевой вывод работы – в БС сформирована совокупность корпоративных знаний, представленных в виде различных категорий данных, на основе которых осуществляется обучение доменно-специализированных языковых моделей [17], основанная на архитектуре-трансформер, предназначенная для предобучения языковых представлений с целью их последующего применения в широком спектре задач обработки естественного языка.

DSLМ, интегрированные в архитектуру ИИ-агентов, используются для автоматизации процессов и снижения объема рутинных операций человека. Качество данных, применяемых при обучении и эксплуатации DSLМ, напрямую определяет устойчивость фокуса моделей и эффективность действий ИИ-агентов, что подтверждает необходимость системного управления качеством данных в интеллектуально-ориентированных корпоративных платформах.

**Рецензент: Сухов Андрей Владимирович, доктор технических наук, профессор, старший научный сотрудник научно-производственного объединения «Специальная техника и связь», г. Москва, Российская Федерация.
E-mail: avs57@mail.ru**

Список литературы / References

1. Бурый А.С., Лопатин И.Н., Погодин И.М. Стандартизация данных и процессов в открытой информационной среде. Часть 1. Открытые данные // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 6(87). С. 469–474. – URL: <https://doi.org/10.24412/2311-1348-2025-6-469-474>. / Buryi A.S., Lopatin I.N., Pogodin I.M. Standartizatsiya dannyh i processov v otkrytoj informacionnoj srede. Part 1. Otkrytye dannye. Information and economic aspects of standardization and technical regulation. 2025;6(87):469–474. (In Russ.).
2. Бурый А.С., Ловцов Д.А. Открытая наука как движущая сила информационного общества // Правовая информатика. 2024. № 3. С. 4–12. – URL: <https://doi.org/10.24682/1994-1404-2024-3-4-12>. / Buryi A.S., Lovtsov D.A. Otkrytaya nauka kak dvizhushchaya sila informacionnogo obshchestva. Legal informatics. 2024;3:4–12. (In Russ.).
3. Дубинин С.К., Теличко Л.Е. Банковская экосистема как клиентоориентированная бизнес-модель в условиях цифровизации // Финансы и кредит. 2022. Т. 28, № 5. С. 1000–1026. – URL: <https://doi.org/10.24891/fc.28.5.1000>. / Dubinin S.K., Telichko L.E. Bankovskaya ekosistema kak klientoorientirovannaya biznes-model v usloviyah cifrovizatsii. Finance and Credit. 2022;28(5):1000–1026. (In Russ.).
4. Васильева Е.В., Солянов К.С., Коневцева Т.Д. Адаптивное хранилище данных как технологический базис экосистемы банка // Финансы: теория и практика. 2020. Т. 24, № 3. С. 132–146. / Vasileva E.V., Solyanov K.S., Konevtseva T.D. Adaptivnoe hranilishche dannyh kak tekhnologicheskij bazis ekosistemy banka. Finansy: teoriya i praktika. 2020;24(3):132–146. (In Russ.).
5. Zhao X., Lu J., Deng C., et al. Beyond one-model-fits-all: A survey of domain specialization for large language models. arXiv preprint arXiv. 2023, 2305.
6. Parate S., Reddi L.T., Agarwal S., Suryadevara M. Analyzing the impact of open data ecosystems and standardized interfaces on product development and innovation. International Journal of Advanced Research in Science, Communication and Technology. 2023;3(1):476–485.
7. Лопатин И.Н. Многоуровневые системы качественных данных на основе моделей искусственного интеллекта: проблемы и решения // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 1(82). С. 70–75. / Lopatin I.N. Mnogourovnevye sistemy kachestvennyh dannyh na osnove modelej iskusstvennogo intellekta: problemy i resheniya. Information and economic aspects of standardization and technical regulation. 2025;1(82):70–75. (In Russ.).
8. Kumar T.V. Cloud-Based Core Banking Systems Using Microservices Architecture. IJRECE. 2019;7(2): 3663–3672.

9. Бурый А.С. Тенденции развития распределенных информационных систем на основе облачных технологий // Транспортное дело России. 2013. № 6. С. 160–162. / Buryi A.S. Tendencii razvitiya raspredelennykh informacionnykh sistem na osnove oblachnykh tekhnologij. *Transportnoe delo Rossii*. 2013;6:160–162. (In Russ.).
10. Бурый А.С., Погодин И.М. Оценка качества больших данных. Часть 2. Модели данных // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 4 (79). С. 24–32. / Buryi A.S., Pogodin I.M. *Ocenka kachestva bol'shih dannyh. Part 2. Data Models. Information and economic aspects of standardization and technical regulation*. 2024;3(78):49–58. (In Russ.).
11. Balaskas G., Papadopoulos H., Pappa D., et al. A Framework for Domain-Specific Dataset Creation and Adaptation of Large Language Models. *Computers*. 2025;14(5): 172.
12. Pessanha Santos N. The expansion of data science: dataset standardization. *Standards*. 2023; 3(4):400–410.
13. Решетников В.Н., Болодурина И.П., Парфенов Д.И. Моделирование размещения сервис-ориентированных приложений в программно-управляемой инфраструктуре виртуального центра обработки данных // Программные продукты и системы. 2016. № 4. С. 15–22. / Reshetnikov V.N., Bolodurina I.P., Parfenov D.I. *Modelirovanie razmeshcheniya servis-orientirovannykh prilozhenij v programmno-upravlyaeмой infrastrukture virtual'nogo centra obrabotki dannyh. Programmnyye produkty i sistemy*. 2016;4:15–22. (In Russ.).
14. Blohm I., Wortmann F., et al. Data products, data mesh, and data fabric: New paradigm (s) for data and analytics? *Business & Information Systems Engineering*. 2024;66(5):643–652.
15. Бурый А.С., Фролов В.А., Куляница А.Л. Эволюция агентного моделирования. Часть 1. Архитектура интеллектуального агента // Информационно-экономические аспекты стандартизации и технического регулирования. 2023. № 5(74). С. 38–47. / Buryi A.S., Frolov V.A., Kulaynitsa A.L. *Evolyuciya agentnogo modelirovaniya. Part 1. Arhitektura intellektualnogo agenta. Information and economic aspects of standardization and technical regulation*. 2023;5(74):38–47. (In Russ.).
16. Jeong C. Fine-tuning and Utilization Methods of Domain-specific LLMs. arXiv 2024. arXiv preprint:2401.02981.
17. Rostam Z.R.K., Kertesz G. Fine-tuning large language models for scientific text classification: A comparative study. In 2024 IEEE 6th International Symposium on Logistics and Industrial Informatics (LINDI). IEEE, 2024: 000233-000238.