

<https://doi.org/10.29188/2712-9217-2023-9-4-7-13>

Искусственный интеллект в здравоохранении России: сбор и подготовка данных для машинного обучения

Оригинальное исследование

А.М. Ханов¹, А.В. Гусев², А.Г. Тюрганов³

¹ ООО «Медицинский аудит, сервис и консалтинг»; д. 31/1, Бирский тракт, Уфа, 450045, Россия

² ООО «К-Скай»; дом 17, наб. Варкауса, Петрозаводск, 185031, Россия

³ ИП НТЦ «Семантика»; дом 99, ул. Достоевского, Уфа, 450005, Россия

Контакт: Ханов Айрат Мидхатович, khanov.a@mail.ru

Аннотация:

Введение. Данная научная статья обращает внимание на важность сбора и подготовки качественных медицинских данных для развития систем искусственного интеллекта (ИИ) в российском здравоохранении.

Материалы и методы. Основной акцент делается на необходимости создания единого федерального стандарта сбора структурированных оцифрованных медицинских данных, позволяющего унифицировать процессы сбора и формирования датасетов для медицинских организаций.

Результаты. В статье предлагается внедрение цифровых помощников, включая доврачебные диагностические опросники, мобильные приложения и программные модули для врачебного осмотра, что снизит затраты времени на заполнение медицинской документации и обеспечит сбор более полной и точной информации. Следующий шаг – реинжиниринг процесса сбора медицинских данных, включая участие пациентов через смартфоны и персональные медицинские помощники. Статья выявляет методологические проблемы, такие как недостоверность, неполнота и недостаток экстенциональных знаний в собранных данных, и предлагает план развития системы сбора медицинских данных, включая создание цифровых стандартов, методов сбора данных и разработку поддержки принятия решений на основе ИИ.

Выводы. Решение описанных проблем и план развития данных имеют важное значение для успешной реализации национального проекта «Экономика данных», поддерживая развитие и применение систем ИИ в медицинской сфере.

Ключевые слова: искусственный интеллект; машинное обучение; медицина; здравоохранение.

Для цитирования: Ханов А.М., Гусев А.В., Тюрганов А.Г. Искусственный интеллект в здравоохранении России: сбор и подготовка данных для машинного обучения. Российский журнал телемедицины и электронного здравоохранения 2023;9(4):7-13; <https://doi.org/10.29188/2712-9217-2023-9-4-7-13>

Artificial intelligence in Russian healthcare: collecting and preparing data for machine learning
Original research

<https://doi.org/10.29188/2712-9217-2023-9-4-7-13>

A.M. Khanov¹, A.V. Gusev², A.G. Tyurganov³

¹ LLC «Medical Audit, Service and Consulting»; 31/1, Birsky Trakt, Ufa, 450045, Russia

² «K-Sky» LLC; build. 17, emb. Varkausa, Petrozavodsk, 185031, Russia

³ IP STC «Semantics»; build. 99, st. Dostoevsky, Ufa, 450005, Russia

Contact: Airat M. Khanov, khanov.a@mail.ru

Abstract:

Introduction. This scientific article draws attention to the importance of collecting and preparing high-quality medical data for the development of artificial intelligence (AI) systems in Russian healthcare.

Materials and methods. The main emphasis is on the need to create a unified federal standard for collecting structured digitized medical data, which will unify the processes of collecting and generating datasets for medical organizations.

Results. The article proposes the introduction of digital assistants, including pre-medical diagnostic questionnaires, mobile applications and software modules for medical examinations, which will reduce the time spent filling out medical documentation and ensure the collection of more complete and accurate information.

The next step is to reengineer the health data collection process, including patient engagement via smartphones and personal health assistants. The article identifies methodological problems such as unreliability, incompleteness and lack of extensional knowledge in the collected data, and proposes a roadmap for the development of a health data collection system, including the creation of digital standards, data collection methods and the development of AI-based decision support.

Conclusions. Solving the described problems and the data development plan are important for the successful implementation of the national project «Data Economy», supporting the development and application of AI systems in the medical field.

Key words: artificial intelligence; machine learning; medicine; healthcare.

For citation: Khanov A.M., Gusev A.V., Tyurganov A.G. Artificial intelligence in Russian healthcare: collecting and preparing data for machine learning. Russian Journal of Telemedicine and E-Health 2023;9(4):7-13; <https://doi.org/10.29188/2712-9217-2023-9-4-7-13>

■ ВВЕДЕНИЕ

Системы искусственного интеллекта (СИИ), применяемые в здравоохранении России, в основном, – это автоматически обучаемые системы (системы машинного обучения). При правильно выбранном математическом аппарате и методе обучения результат обучения всецело зависит от качества обучающей выборки – набора структурированных оцифрованных медицинских данных (датасета). Обеспечить требуемое качество данных возможно только в случае правильной организации процессов сбора и предварительной подготовки данных. Этой проблеме посвящена настоящая работа.

■ МАТЕРИАЛЫ И МЕТОДЫ

Основной акцент делается на необходимости создания единого федерального стандарта сбора структурированных оцифрованных медицинских данных, позволяющего унифицировать процессы сбора и формирования датасетов для медицинских организаций.

■ РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Необходимость совершенствования сбора и подготовки данных для машинного обучения в здравоохранении

В отрасли здравоохранения для развития СИИ и систем поддержки принятия врачебных решений (СППВР) в первую очередь необходимы качественные датасеты из структурированных оцифрованных медицинских данных с максимальной полнотой охвата набора признаков распознавания, результирующими классов распознаваемых образов, первичных источников данных, результатов решений задач распознавания экспертами, потоков сбора, обработки и выдачи информации.

В этой связи должна быть выстроена логическая цепочка: структурированные оцифрованные медицинские данные – качественные датасеты – ИИ (СППВР) (рис. 1). Сейчас у нас существенный ряд проблем в первом, а соответственно и втором этапах, являющихся фундаментом всего.

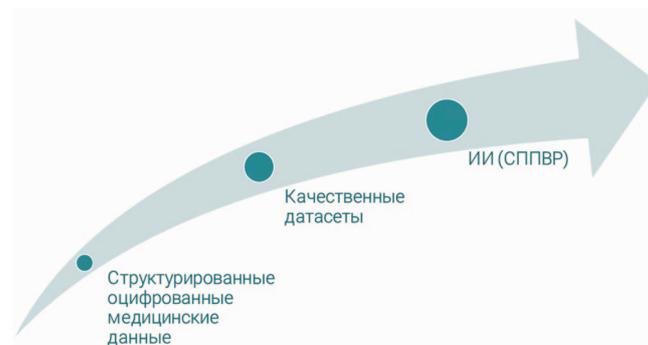


Рис. 1. Этапность развития СИИ для здравоохранения
Fig. 1. Stages of development of AI for healthcare

Данные для принятия организационно-управленческих решений как в отдельных медицинских организациях, так и по отрасли в целом (медицинские транзакции, факты выполненных работ и расходования ресурсов) оцифрованы и присутствуют во всех медицинских информационных системах (МИС). А данные для врачебных решений в ходе процесса диагностики, маршрутизации и лечения пациентов (жалобы, анамнез, результаты осмотра врача) оцифрованы частично, присутствуют в МИС в основном в текстовом неструктурированном формате, включая:

- МКБ–Х – оцифровано, предстоит переход на МКБ-ХI;
- Лабораторные данные – в основном, цифровые, автоматически получаются из лабораторного оборудования;
- Данные диагностических исследований – частично, только из цифровых диагностических систем;
- Морфология – оцифровка в начале пути;
- Описание осмотра врача – частично структурировано, в большей части текстовый формат;
- Жалобы и анамнез пациента – текстовый формат. В настоящее время практически все, что говорит пациент и видит врач на амбулаторном приеме вводится в медицинские информационные системы в сокращенном текстовом формате и не может непосредственно без потерь и искажений использоваться для цифровой обработки. Необходим полноценный структурированный цифровой медицинский профиль пациента;

- Клинические рекомендации – оцифрованы частично.

2. Единый Федеральный стандарт сбора медицинских данных

Для развития и массового применения СИИ в здравоохранении РФ важно построить систему сбора структурированных и оцифрованных медицинских данных, начиная с анамнеза, жалоб и данных осмотра врача. При этом требуется строить ее не в рамках одной МИС для одной или нескольких МО или регионов, а создать и внедрить единый федеральный стандарт (формат данных, протокол информационного обмена и регламент сбора) для всех программно-информационных и программно-аппаратных медицинских систем страны (МИС МО и других). Тогда становится возможной унификация процессов сбора данных и формирования качественных датасетов для любой медицинской организации в строящемся сейчас домене «Здравоохранение». И вооруженные системами искусственного интеллекта врачи всех регионов смогут читать и использовать их.

В настоящее время получение качественных наборов данных для СИИ является крайне сложной задачей. Как правило, создание наборов данных осуществляется ретроспективно и на основе данных реальной клинической практики (ДРКП, RWD). На рисунке 2 зеленым отмечены структурированные области данных и желтым ►►

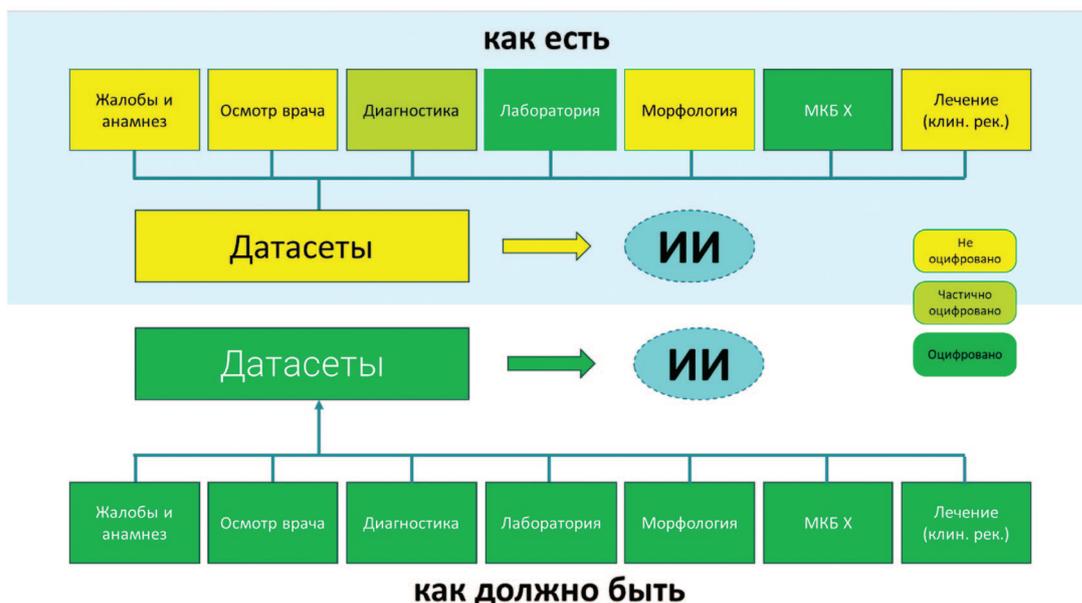


Рис. 2. Информационные потоки при формировании наборов данных для машинного обучения
Fig. 2. Information flows when generating data sets for machine learning

не оцифрованные. Датасеты, где собрана информация о лечении, очень часто не соответствующим клиническим рекомендациям, к качественной работе систем ИИ не приведут.

3. Реинжиниринг сбора медицинских данных

При разработке и внедрении комплексной стандартизированной системы структурирования и оцифровки медицинских данных, наборы данных необходимо формировать **проспективно** с первого дня внедрения, последовательно и полноценно. При широком внедрении процесс сбора необходимых и достаточных для обучения СИИ объемов данных может быть очень быстрым. На таких базах данных со временем можно будет строить и многофункциональные системы ИИ.

При этом появляется еще одна важная задача: **реинжиниринг сбора медицинских данных** [1]. Сейчас это делает врач, который является единственным источником данных о здоровье пациента в МИС. Только врач своими руками вводит все данные в МИС, что требует больших затрат времени и приводит к низкому качеству собираемой информации. Необходимо привлечь к этому процессу население – пациентов через имеющийся практически у всех гаджеты, персональные медицинские помощники, носимые устройства и т.п. (рис. 3).

Практически у всего населения есть смартфоны, которые должны стать цифровым инструментом врача в руках пациента. Например,

рассмотрим идею создания и внедрения доврачебных диагностических опросников [2] в мобильном приложении, программе на компьютере или терминале в поликлинике.

В результате внедрения приложения с доврачебными диагностическими опросниками можно ожидать следующих эффектов:

- Для врача:

- о снижение нагрузки и оптимизация приема за счет сокращения на 3-5 минут (порядка 30%) времени опроса и ввода данных в МИС;

- о обеспечение необходимой полноты информации для принятия врачебных решений;

- о предоставление возможности дистанционного получения данных пациента;

- о обеспечение средствами удаленного мониторинга пациентов в режиме 24/7 с обратной связью.

- Для пациента:

- о возможность быстрого и удобного обращения к врачу;

- о безопасное хранение своих медицинских данных в облаке;

- о предоставление доступа к данным медицинским работникам;

- о удаленная эффективная коммуникация с врачом;

- о возможность даже при амбулаторном лечении находиться под постоянным контролем медиков.

Единый состав доврачебных диагностических опросников в мобильных приложениях для



Рис. 3. Сбор медицинских данных от пациентов с помощью персональных медицинских помощников
Fig. 3. Collecting medical data from patients using personal medical assistants

различных МИС может позволить использовать эту информацию в федеральном масштабе (рис. 4).

Следующее средство автоматизации сбора структурированных данных – приложение «Врачебный осмотр». Такой стандартизованный модуль для любой МИС домена «Здравоохранение» значительно снизит издержки врача на заполнение первичной медицинской докумен-

тации. Программное обеспечение позволит практически исключить работу с клавиатурой, основными инструментами ввода данных сделать сенсорную панель, мышь и микрофон.

Предлагаемый подход формирует семейство цифровых помощников, автоматизирующих процесс сбора первичных актуальных структурированных медицинских данных на всех этапах амбулаторной медицинской помощи (табл. 1). ►►



Рис. 4. Схема сбора структурированных данных с помощью программных приложений «Доврачебный опросник» и «Врачебный осмотр»
Fig. 4. Scheme for collecting structured data using the software applications «Pre-medical Questionnaire» and «Medical Examination»

Таблица 1. Цифровая поддержка амбулаторной медицинской помощи
Table 1. Digital support for outpatient medical care

ЭТАПЫ АМП	ЦИФРОВАЯ ПОДДЕРЖКА	ПАЦИЕНТ	ВРАЧ	ЦИФРОВАЯ ПОДДЕРЖКА
Подготовка визита к врачу	Доврачебный диагностический опросник в мобильном приложении/терминале в регистратуре	Заполнение доврачебного диагностического опросника по профилю врача		СППВР – Диагностические гипотезы – 1 (на базе результатов доврачебного диагностического опросника)
Прием врача первичный		Визит к врачу	<ul style="list-style-type: none"> • Врачебный осмотр • Установка предварительного диагноза • Программа дообследования 	Цифровой помощник – «Врачебный осмотр» СППВР – Диагностические гипотезы – 2 (+ данные врачебного осмотра) СППВР – Клинические рекомендации по программе диагностики (выбор из возможных вариантов)
Дообследование		Диагностические процедуры		Системы ИИ
Прием врача повторный		Визит к врачу	<ul style="list-style-type: none"> • Установка диагноза • Назначение лечения 	СППВР – Диагностические гипотезы – 3 (+ лабораторные и диагностические данные). СППВР – Клинические рекомендации по протоколам лечения (выбор из возможных вариантов)
Амбулаторное лечение	Мониторинговый опросник в мобильном приложении	Амбулаторное лечение	Удаленный мониторинг	Мониторинг состояния через асинхронную коммуникацию
Прием врача повторный		Визит к врачу	<ul style="list-style-type: none"> • Контроль эффективности лечения • Рекомендации 	Результаты мониторинга
Диспансерное наблюдение	Мониторинговый опросник в мобильном приложении	Удаленное наблюдение	Удаленный мониторинг	Мониторинг состояния через асинхронную коммуникацию
Диспансеризация – первый этап	Доврачебный диагностический опросник в мобильном приложении	Ежегодное заполнение диагностического опросника по программе диспансеризации		Система ИИ Маршрутизация на второй этап

4. Методологические проблемы сбора медицинских данных

Проблемы сбора данных для построения систем искусственного интеллекта в отечественном здравоохранении с методологической точки зрения выглядят следующим образом:

- при формировании датасетов, как правило, предварительно не используется формализация постановки задачи сбора данных, гипотез ожидаемых зависимостей и т. п.
- не все необходимые для решения задач данные собираются (неполнота и нарушение целостности данных, непредставительность данных);
- как правило, для обучения СИИ собираются ретроспективные неактуальные противоречивые данные, которые невозможно верифицировать (недоверенность и противоречивость данных);
- при сборе данных не все источники рассматриваются к использованию (как правило, данные поступают только от врача и в интерпретации врача, нет данных непосредственно от пациента, ЕГИСЗ замкнута относительно источников сбора данных, нет возможности подключить новые);
- для обучения СИИ нужны не просто данные, а экстенциональные знания, пригодные для автоматического применения соответствующими программными средствами;
- для формирования качественных датасетов необходимо проектировать модели и структуры данных, а также модели информационных потоков, модели сбора, представления и обработки данных в процессе обучения и применения СИИ, модели адекватности и формирования погрешности [3].

Описанные методологические проблемы сбора данных существенно сдерживают развитие СИИ в медицине и требуют основательного системного подхода при решении задач формирования медицинских датасетов.

5. План развития системы сбора медицинских данных

Учитывая представленные выше соображения, нами предлагается следующий план организации сбора и подготовки данных для развития ИИ в здравоохранении России:

- Разработка единого цифрового стандарта сбора медицинских данных:
 - о Создание и валидация профильных до-врачебных диагностических и мониторинговых опросников (для пациентов).
 - о Формирование цифрового справочника для описания результата осмотра врача по системам и органам (для врачей).
 - о Цифровизация клинических рекомендаций (диагностика/лечение) – структурированный список всех возможных вариантов, прописанных в утвержденных МЗ РФ клинических рекомендациях. Для каждого пациента – структурированный «отфильтрованный» список, за исключением вариантов, противопоказанных наличием определенных оцифрованных медицинских данных (жалобы, анамнез, осмотр, диагностика, динамика).
- Разработка методов сбора данных:
 - о Разработка и внедрение инструмента асинхронной цифровой коммуникации пациента с медицинскими организациями и врачами (мобильное приложение/терминал).
 - о Разработка и внедрение универсального (для всех МИС) цифрового модуля «Врачебный осмотр» – единой справочной системы описания состояния здоровья с поддержкой звуковым чат-ботом.
- Реинжиниринг сбора первичных медицинских данных – сбор данных с участием пациента (доврачебный диагностический опрос пациента в мобильном приложении, терминал в регистратуре).
- Создание цифрового портрета (профиля) пациента на основе формализованных данных и знаний, в статике и динамике.
- Разработка систем поддержки принятия врачебных и пациентских решений в виде обучающихся и экспертных систем.
- Разработка систем искусственного интеллекта.
- Разработка моделей, методов и программно-информационных инструментальных средств управления диагностикой, маршрутизацией и лечением пациентов на основе данных.

■ ВЫВОДЫ

Объявление Президентом России о формировании национального проекта «Экономика

данных», первым пунктом которого выделена задача сбора данных, подчеркивает важность решения описанных в настоящей статье про-

блем. Предлагаем сосредоточить усилия специалистов в области медицинской информатики в этом направлении. //

ЛИТЕРАТУРА

1. Павлов В.Н., Ханов А.М., Тюрганов А.Г. Цифровая медицина: ожидания и перспективы. *Общественное здоровье* 2022;2(2):73-6. [Pavlov V.N., Khanov A.M., Tyurganov A.G. Digital medicine: expectations and prospects. *Obshchestvennoye zdorov'ye = Public Health* 2022;2(2):73-6. (In Russian)]. <https://doi.org/10.21045/2782-1676-2022-2-2-73-76>.
2. Буш Е. От «истории болезни» к «истории здоровья». Управление на основе данных – ключ к цифровой трансформации отрасли. *Медицинская газета* 2022;49:10. [Bush E. From «medical history» to «health history». Data-driven management is the key to digital transformation of the industry. *Meditsinskaya gazeta = Med Gazette* 2022;49:10. (In Russian)].
3. Зверев Г.Н. Теоретическая информатика и ее основания. *Физматлит* 2008;2:576. [Zverev G.N. Theoretical computer science and its foundations. *Fizmatlit* 2008;2:576. (In Russian)].
4. Пугачев П.С., Гусев А.В., Кобякова О.С., Кадыров Ф.Н., Гаврилов Д.В., Новицкий Р.Э., Владимирский А.В. Мировые тренды цифровой трансформации отрасли здравоохранения. *Национальное здравоохранение* 2021;2(2):5-12. [Pugachev P.S., Gusev A.V., Kobyakova O.S., Kadyrov F.N., Gavrilov D.V., Novitsky R.E., Vladimirsky A.V. Global trends in digital transformation of the healthcare industry. *Natsional'noye zdravookhraneniye = National Health* 2021;2(2):5-12 (In Russian)]. <https://doi.org/10.47093/2713-069X.2021.2.2.5-12>.
5. Гусев А.В., Владимирский А.В., Голубев Н.А., Зарубина Т.В. Информатизация здравоохранения Российской Федерации: история и результаты развития. *Национальное здравоохранение* 2021;2(3):5–17. [Gusev A.V., Vladimirsky A.V., Golubev N.A., Zarubina T.V. Informatization of healthcare in the Russian Federation: history and development results. *Natsional'noye zdravookhraneniye = National Health* 2021;2(3):5–17. (In Russian)].

<https://doi.org/10.47093/2713-069X.2021.2.3.5-17>.

6. Карпов О.Э., Храмов А.Е. Информационные технологии, вычислительные системы и искусственный интеллект в медицине. *ДПК Пресс* 2022:480. [Karpov O.E., Khramov A.E. Information technologies, computing systems and artificial intelligence in medicine. *DPK Press* 2022:480. (In Russian)]. <https://doi.org/10.56463/krphrm-978-5-91976-232-4>.
7. Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). [Электронный ресурс]. [Decree of the President of the Russian Federation dated October 10, 2019 No. 490 «On the development of artificial intelligence in the Russian Federation» (together with the «National Strategy for the Development of Artificial Intelligence for the period until 2030»). [Electronic resource]. (In Russian)]. URL: http://www.consultant.ru/document/cons_doc_LAW_335184/.
8. Гусев А.В., Реброва О.Ю. Осведомленность и мнения руководителей в сфере здравоохранения России о медицинских технологиях искусственного интеллекта. *Врач и информационные технологии* 2023;1(4):28-39. [Gusev A.V., Rebrova O.Yu. Awareness and opinions of managers in the Russian healthcare sector about medical technologies of artificial intelligence. *Vrach i informatsionnyye tekhnologii 2 = Physician and Information Technology* 2023;1(4):28-39. (In Russian)]. https://doi.org/110.25881/18110193_2023_1_28.
9. Orlova I.A., Akopyan Zh.A., Plisyuk A.G., Tarasova E.V., Borisov E.N., Dolgushin G.O., et al. Opinion research among Russian Physicians on the application of technologies using artificial intelligence in the field of medicine and health care. *BMC Health Services Research* 2023;23:749. <https://doi.org/10.1186/s12913-023-09493-6>.

Сведения об авторах:

Ханов А.М. – д.м.н., профессор, ООО «Медицинский аудит, сервис и консалтинг»; Уфа, Россия; RINЦ Author ID 881342

Гусев А.В. – к.т.н., директор по развитию ООО «К-Скай»; Петрозаводск, Россия; RINЦ Author ID 168742

Тюрганов А.Г. – к.т.н., доцент, ИП НТЦ «Семантика»; Уфа, Россия

Вклад авторов:

Ханов А.М. – определение научного интереса и актуальности тематики, обзор литературы, 40%
Гусев А.В. – литературный обзор, написание текста статьи, 30%
Тюрганов А.Г. – определение научного интереса и актуальности тематики, написание текста статьи, 30%

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов.

Финансирование: Исследование проведено без спонсорской поддержки.

Статья поступила: 10.10.2023

Рецензирование: 12.11.2023

Исправления получены: 18.11.23

Принята к публикации: 23.12.2023

Information about authors:

Khanov A.M. – Dr. Sci., Professor, Medical Audit, Service and Consulting LLC; Ufa, Russia; RSCI Author ID 881342

Gusev A.V. – PhD, Development Director of K-Sky LLC; Petrozavodsk, Russia; RSCI Author ID 168742

Tyurganov A.G. – PhD, Associate Professor, IP STC «Semantics»; Ufa, Russia

Authors contributions:

Khanov A.M. – determination of scientific interest and relevance of the topic, literature review, 40%
Gusev A.V. – literature review, writing the text of the article, 30%
Tyurganov A.G. – determination of scientific interest and relevance of the topic, writing the text of the article, 30%

Conflict of interest. The author declare no conflict of interest.

Financing. The study was performed without external funding.

Received: 10.10.2023

Reviewing: 12.11.2023

Reviewing: 18.11.2023

Accepted for publication: 23.12.2023