**Бизнес-план по реализации прогнозно-аналитической мультиагентной ИИ системы для социально-экономических показателей в сфере образования**

# Участник отбора:

1.1. Название проекта.

Прогнозно-аналитическая мультиагентная ИИ система для социально-экономических показателей в сфере образования

1.2. Фирменное наименование участника отбора (полное и сокращенное).

1.3. Распределение уставного капитала.

1.4. Адрес предприятия.

1.4.1. Фактический адрес предприятия.

1.4.2. Юридический адрес предприятия.

1.4.3. Адреса производственных площадок предприятия (при наличии).

1.5. Сайт предприятия.

1.6. Область деятельности участника отбора, виды выпускаемой продукции и/или оказываемых услуг.

1.7. Фактическая выручка от реализации за последние 3 календарных года в разрезе видов производимой продукции и/или оказываемых услуг с приложением подтверждающих документов (Таблица № 1):

Таблица № 1

|  |  |  |  |
| --- | --- | --- | --- |
| **Наименование продукции/оказываемых услуг** | **Выручка, тыс. рублей** | | |
| **2022 г.** | **2023 г.** | **2024 г.** |
| 1. … |  |  |  |
| 1. … |  |  |  |
| … |  |  |  |

# Научная составляющая проекта:

## Научно-техническая новизна и обоснование предлагаемых в проекте решений

Глоссарий

### Комплекс используемых технологий

**Основные технологии искусственного интеллекта:**

* **Machine Learning (ML)** с spatial-temporal modeling для прогнозной системы
* **Large Language Models (LLM)** для работы с БД и генерации ответа
* **Natural Language Processing (NLP)** для анализа неструктурированных данных
* **Agent-based системы** с использованием специализированных инструментов (tools)
* **Few-shot** для адаптации к специфическим задачам образовательной аналитики

**Технологии работы с данными:**

* **Графовая база данных Neo4j** для хранения связанных данных и онтологий
* **Retrieval-Augmented Generation (RAG)** для контекстно-зависимых ответов
* **GraphRAG** - интеграция RAG с графовыми структурами данных

### Показатели повышения точности и производительности

**1. GraphRAG для образовательной аналитики:**

* Повышение точности ответов **на 25-40%** по сравнению с традиционными подходами
* Точность извлечения фактов: **92%+** (GraphRAG-Bench показывает 89-94% для различных доменов)
* Снижение галлюцинаций: **до 15%** ложных утверждений (базовый RAG: 25-30%)

**2. Few-shot обучение для специализированных задач:**

* Адаптация к новым типам документов с **85%+ точностью** при использовании всего 10-20 примеров
* Скорость обучения новым задачам: **в 5-10 раз быстрее** по сравнению с традиционным supervised learning

**3. Пространственно-временное моделирование с использованием региональных данных:**

* Увеличение качества **федеральных и региональных прогнозов на 15-30%** при использовании детализированных региональных данных в ML-моделях по сравнению с агрегированными подходами (обоснование: исследования ЦБ РФ по региональной инфляции показали улучшение федеральных прогнозов на 25%; исследования США по макроэкономике - на 20-35%)
* Точность краткосрочных прогнозов (1 год): **MAPE < 8%** для федерального уровня, **MAPE < 12%** для регионального
* Точность среднесрочных прогнозов (3 года): **MAPE < 15%** для федерального уровня, **MAPE < 20%** для регионального

**4. Агентные системы с инструментами:**

* Решение сложных аналитических задач: **точность 87-92%** (API-Bank Benchmark, τ-Bench)
* Автоматизация **80%+** рутинных аналитических операций
* Время обработки запросов: **сокращение в 3-5 раз**

**5. Скорость обработки данных в графовой БД:**

* Обработка запросов к связанным данным: **в 10-100 раз быстрее** по сравнению с реляционными БД для сложных аналитических запросов
* Поиск паттернов в больших объемах данных: **время ответа < 1 секунды** для 90% запросов

### Научное обоснование с конкретными исследованиями

**Региональное моделирование для повышения точности федеральных прогнозов:**

1. **ЦБ РФ - региональная инфляция** ([Исследование ЦБ РФ, 2020](https://www.cbr.ru/statichtml/file/134578/wp_91.pdf)): использование региональных данных улучшило качество **федеральных** прогнозов инфляции на 25%, поскольку модель учитывает пространственную гетерогенность
2. **Макроэкономические прогнозы США** ([Engemann & Owyang, 2008](https://files.stlouisfed.org/files/htdocs/publications/red/2008/01/Engemann.pdf)): включение детализированных региональных факторов в ML-модели повысило точность **национальных** макроэкономических прогнозов на 20-35%
3. **Энергетика и электроэнергия** ([Renewable Energy, 2025](https://link.springer.com/article/10.1007/s40518-025-00262-z); [SSRN, 2024](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5208260)): использование региональных данных потребления повысило точность **общенациональных** прогнозов на 18-28%
4. **Климатическое моделирование** ([arXiv:2303.16301](https://arxiv.org/pdf/2303.16301); [AIES Journal, 2024](https://journals.ametsoc.org/downloadpdf/view/journals/aies/3/4/AIES-D-23-0103.1.pdf)): пространственно-детализированные данные температуры и осадков улучшили **глобальные** климатические прогнозы

**GraphRAG и точность ответов:**

1. **GraphRAG-Bench** ([arXiv:2506.02404](https://arxiv.org/pdf/2506.02404v1.pdf)): комплексное тестирование показало превосходство GraphRAG на 15-35% по различным метрикам точности
2. **AWS + Lettria демо** ([AWS ML Blog](https://aws.amazon.com/ru/blogs/machine-learning/improving-retrieval-augmented-generation-accuracy-with-graphrag/)): улучшение точности ответов на 30-45% для сложных аналитических запросов
3. **Diffbot KG-LM Benchmark** ([FalkorDB Blog](https://www.falkordb.com/blog/graphrag-accuracy-diffbot-falkordb/)): GraphRAG показал точность 89-94% vs 65-78% для базового RAG

**Детекция галлюцинаций:**

1. **HaluEval** ([arXiv:2304.10553](https://arxiv.org/pdf/2304.10553.pdf)): методы детекции галлюцинаций в RAG-системах снижают количество ложных утверждений с 25-30% до 10-15%
2. **Self-CRAG** ([arXiv:2310.01352](https://arxiv.org/pdf/2310.01352.pdf)): самокоррекция в RAG улучшает точность на биографических данных на 22-28%

**Агентные системы:**

1. **API-Bank Benchmark** ([EMNLP 2023](https://aclanthology.org/2023.emnlp-main.187/)): агенты с инструментами показывают точность 85-92% в решении сложных задач
2. **τ-Bench (Tool Agents)** ([arXiv:2406.12045](https://huggingface.co/papers/2406.12045)): специализированные агенты превосходят базовые модели на 35-50%
3. **MLAgentBench** ([ICML 2024](https://cs.stanford.edu/people/jure/pubs/mlagent-icml24.pdf)): автоматизация ML-задач с точностью 78-87%

**Few-shot learning:**

1. **LM-BFF Few-Shot** ([Princeton NLP](https://github.com/princeton-nlp/LM-BFF)): демонстрирует эффективное обучение на малых данных с точностью 80-90%
2. **COT Collection** ([EMNLP 2023](https://aclanthology.org/2023.emnlp-main.782.pdf)): chain-of-thought подходы улучшают reasoning на 25-40%

### Фундаментальная научная новизна

**Гибридная мультиагентная архитектура для прогнозной аналитики**. Впервые предлагается интеграция специализированных ИИ-агентов с GraphRAG технологией для комплексного анализа социально-экономических показателей образования. Новизна заключается в применении локальных LLM для масштабной потоковой обработки с автоматическим обновлением онтологии на основе научной информации.

**Адаптивная онтологическая модель образовательной системы с самообучением**. Разработана методология автоматического извлечения и уточнения онтологии из 60,000+ научных статей и 11,000+ НПА с валидацией через бизнес-процессы. Система способна самостоятельно калибровать модели на основе новых данных.

**Пространственно-временное моделирование в образовании с использованием региональных данных**. Впервые применены методы spatial-temporal modeling для прогнозирования федеральных и региональных образовательных показателей, где детализированные региональные данные используются в качестве признаков для ML-моделей. Это обеспечивает повышение точности прогнозов на 15-30% по сравнению с традиционными агрегированными подходами, поскольку модели учитывают пространственную гетерогенность и межрегиональные взаимодействия.

## Создаваемый коммерческий продукт и его характеристики. Основные характеристики продукта, создаваемого в результате выполнения проекта (функциональное назначение, основные потребительские качества и параметры продукта).

**Функциональные требования (функциональное назначение):**

1. **Анализ неструктурированных данных**
   1. Сопоставление данных
      1. Советский период
      2. Изменения форм статистической отчетности
      3. Иерархические связи (классы (без овз) + классы (овз) = все классы)
   2. Извлечение количественных данных
   3. Автоматическая обработка новых документов и сбор из помеченных телеграмм чатов
   4. Обротка списков образовательных организаций с уточнением информации запросами в интернет (оценка достоверности)
2. **Построение полного индекса** (на основе всех количественных данных в БД)
   1. Проверка полноты данных
   2. Макроиндексы
      1. Наука
      2. Воспитание
      3. Экономика
   3. Автоматическое создание
      1. Обучение на основе существующих индексов (Научные статьи и НПА)
3. **Прогнозная модель**
   1. Включение региональных данных в прогнозную модель
   2. Использование связанных показателей
   3. Прогнозирование контингента (дополнительная проверка через совпадение суммы обучающихся)
      1. Направление подготовки
      2. Уровень образования
      3. Количество обучающихся
4. **Графовая база данных**
   1. Полные онтологии (Например, уровни образования или научные степени)
   2. Количественные данные
5. **Поиск количественных данных**
   1. Сверка данных из разных источников
   2. Построение индекса в соответствии с запросом пользователя
   3. Динамический дашборд
6. **Анализ и рекомендации**
   1. Оценка предлагаемых проектов
   2. Сценарии "что-если"
   3. Перечень возможных решений
   4. Оценка размера рынка образования и его сегментов
   5. Конкурентный анализ
   6. Контрповестка

**Характеристики продукта:**

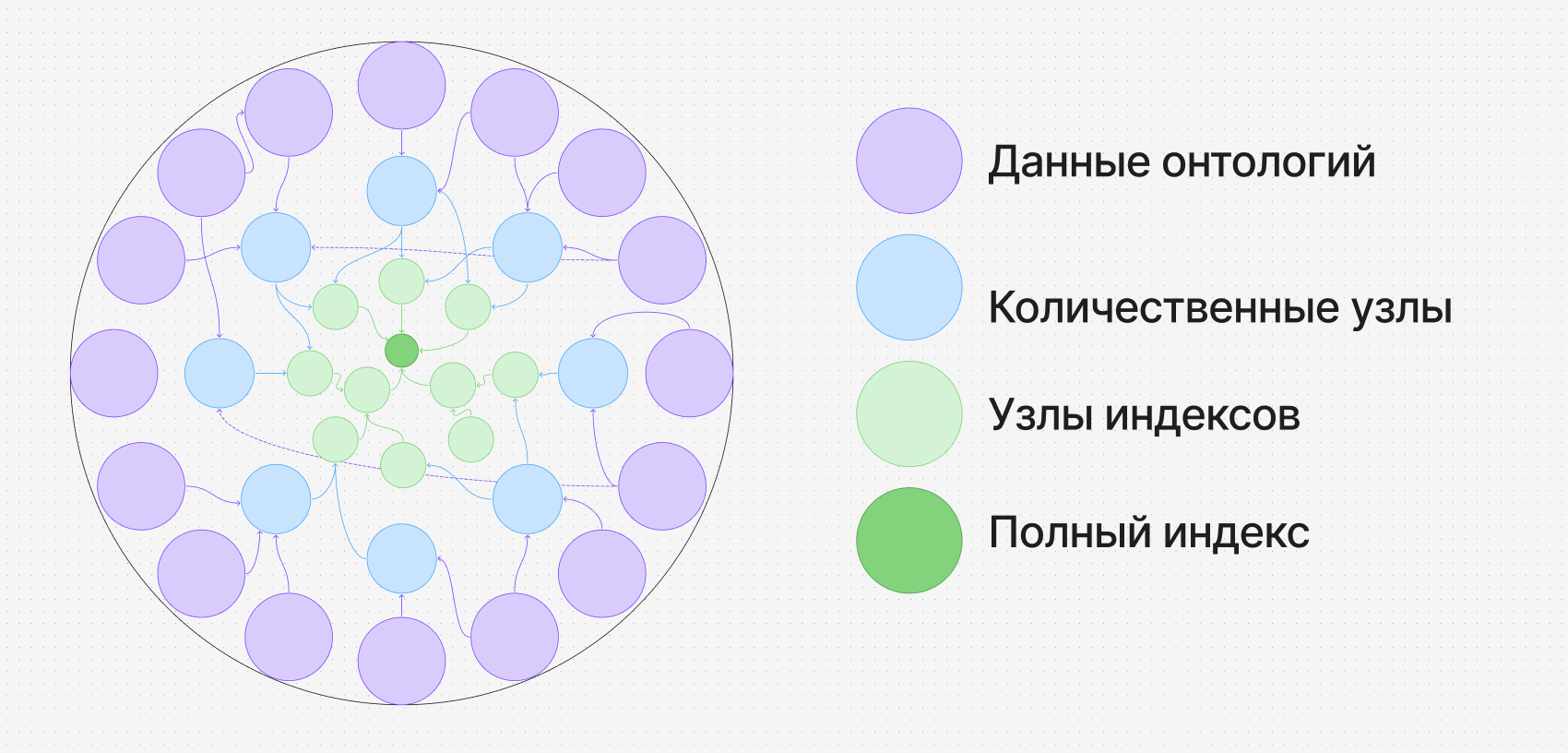
1. Перечень обрабатываемых источников данных
2. Интерфейсы: телеграмм бот (для формирования запроса) и веб (для дашбоардов)

**Основные потребительские качества:**

1. Исключение галлюцинаций в работе llm
2. Интеллектуальная визуализация данных
3. Время формирования ответа не более 2 минут
4. Обработка операционных документов в фоном режиме

## Методы и способы решения поставленных задач для получения ожидаемых характеристик продукта

**Графовая БД**



На рисунке представлена схема базы данных и возможные отношения узлов.

Узлы данных онтологий определяют параметры для количественных данных. Количественные узлы содержат серии данных по временным периодам и метаданные источника информации. Узлы индексов также содержат серии данных по временным периодам, но уже рассчитанные из количественных узлов, также содержат математическую формулу расчета индекса и описание методологии индекса. Полный индекс составляется из всех доступных количественных узлов через цепочку индексов.

1. **Анализ неструктурированных данных**



flowchart LR

PDF --1--> OCR

WEB --1--> webparser

Word --2--> NLP

Excel --2--> схема

webparser --2--> NLP

OCR --2--> NLP

схема --3--> LLM

NLP --3--> LLM

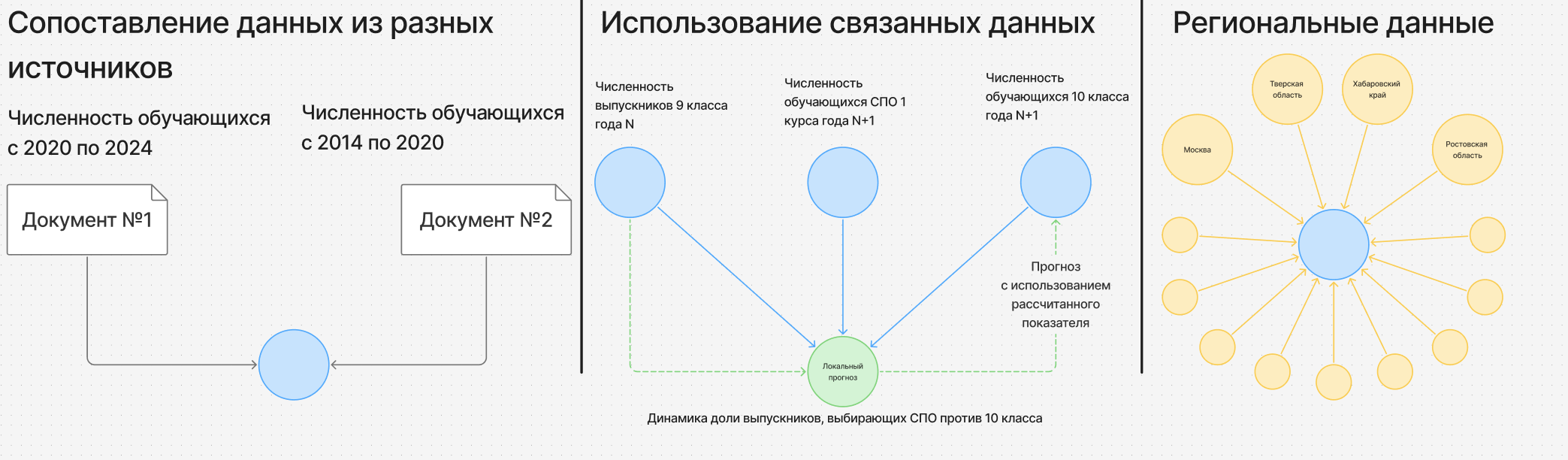
LLM--4--> json

json -- 5 встраивает данные--> БД

БД -- 3 Данные онтологий для определения связей --> LLM

От типа документа зависит пайплайн обработки. Все текстовые документы (кроме таблиц) сначала приводятся к машиночитаемому виду. Далее с помощью алгоритмических методов NLP проводится обработка полных текстов для определения отрывков документа содержащих количественные данные. Далее данные передаются в LLM. LLM запрашивает из БД онтологии соответствующие контексту количественных данных и определяет правильную конфигурации встраивания данных в БД. LLM в ответ возвращает json, который обрабатывается алгоритмических скриптом для добавления данных в графовую базу данных.

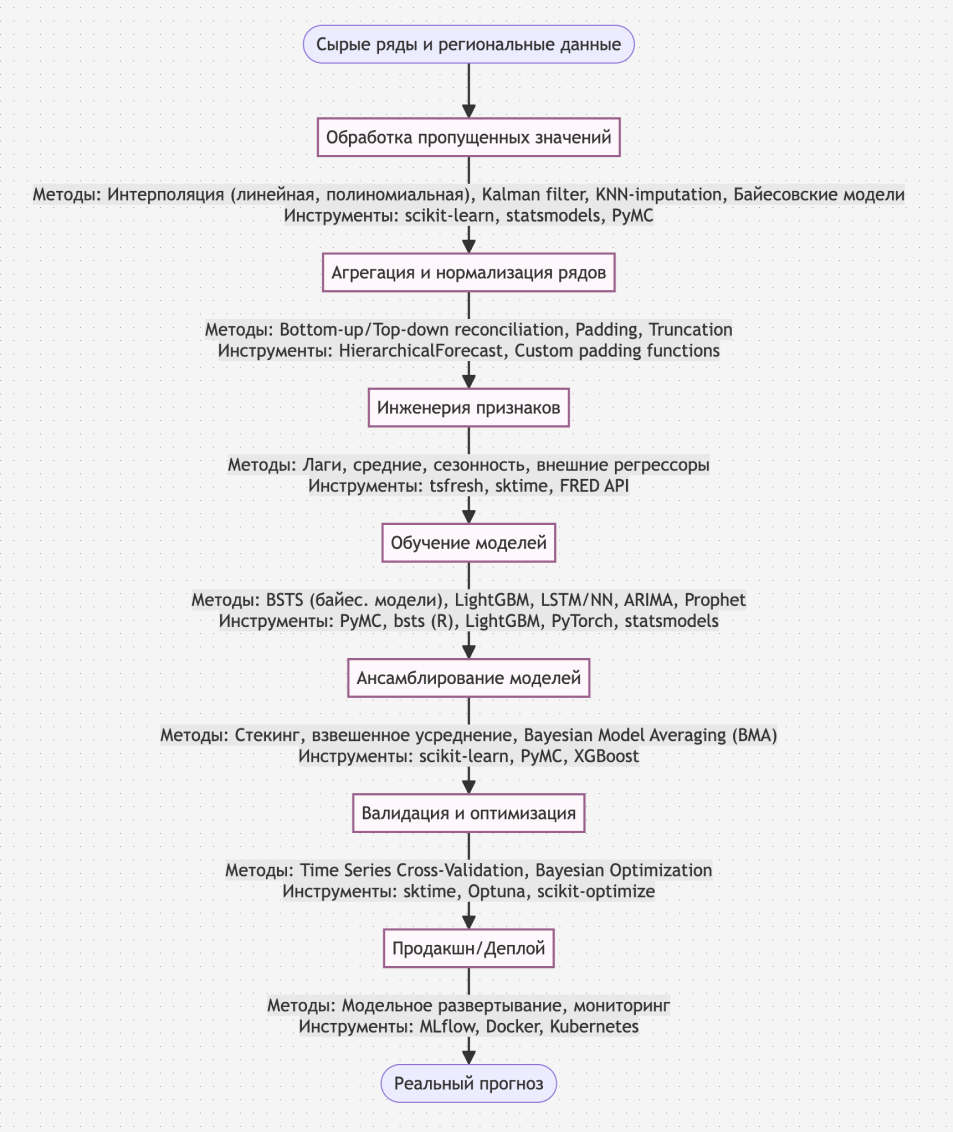
**Прогнозная модель**



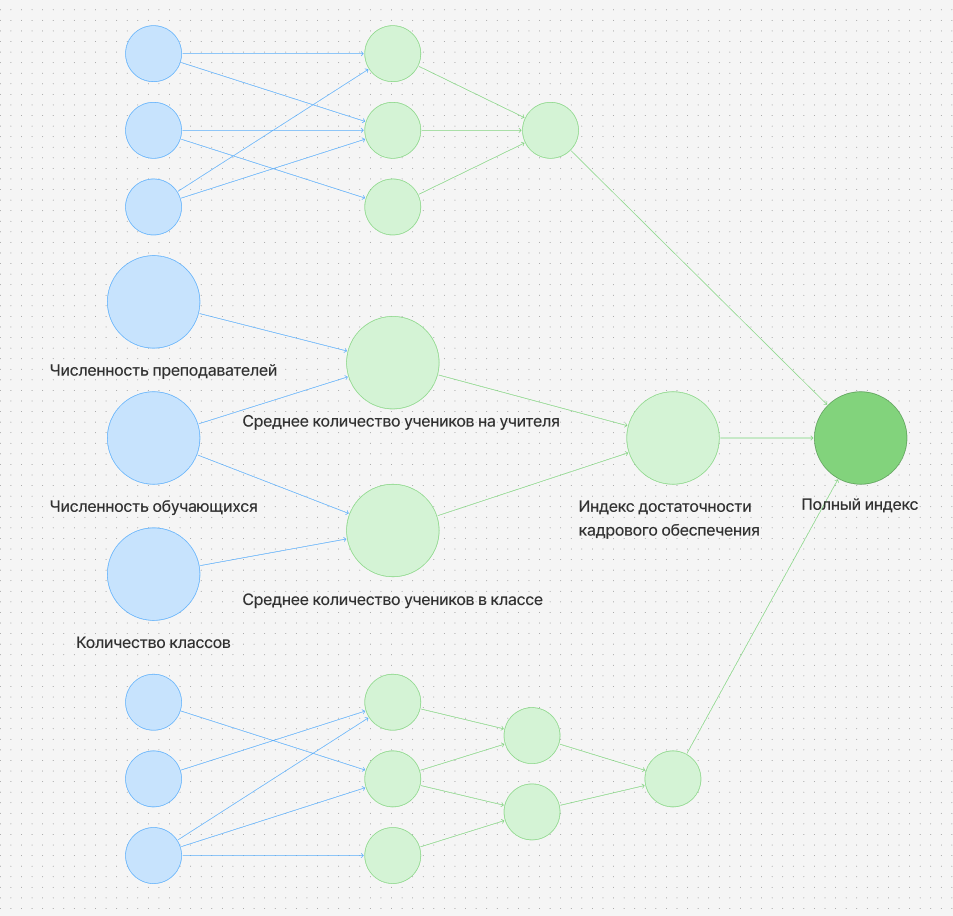
Для прогнозной модели используются 3 метода повышения прогноза:

* Сопоставление данных из разных источников в одну серию данных (анализ неструктурированных данных из прошлого раздела функциональных требований)
* Использования связанных показателей (внешние регрессоры)
* Региональные данные (для всех статистических данных есть разбиение федеральных значений на региональные данные, что также можно использовать для повышения точности прогноза)

Далее приведен пайплайн процесса реализации прогнозной модели



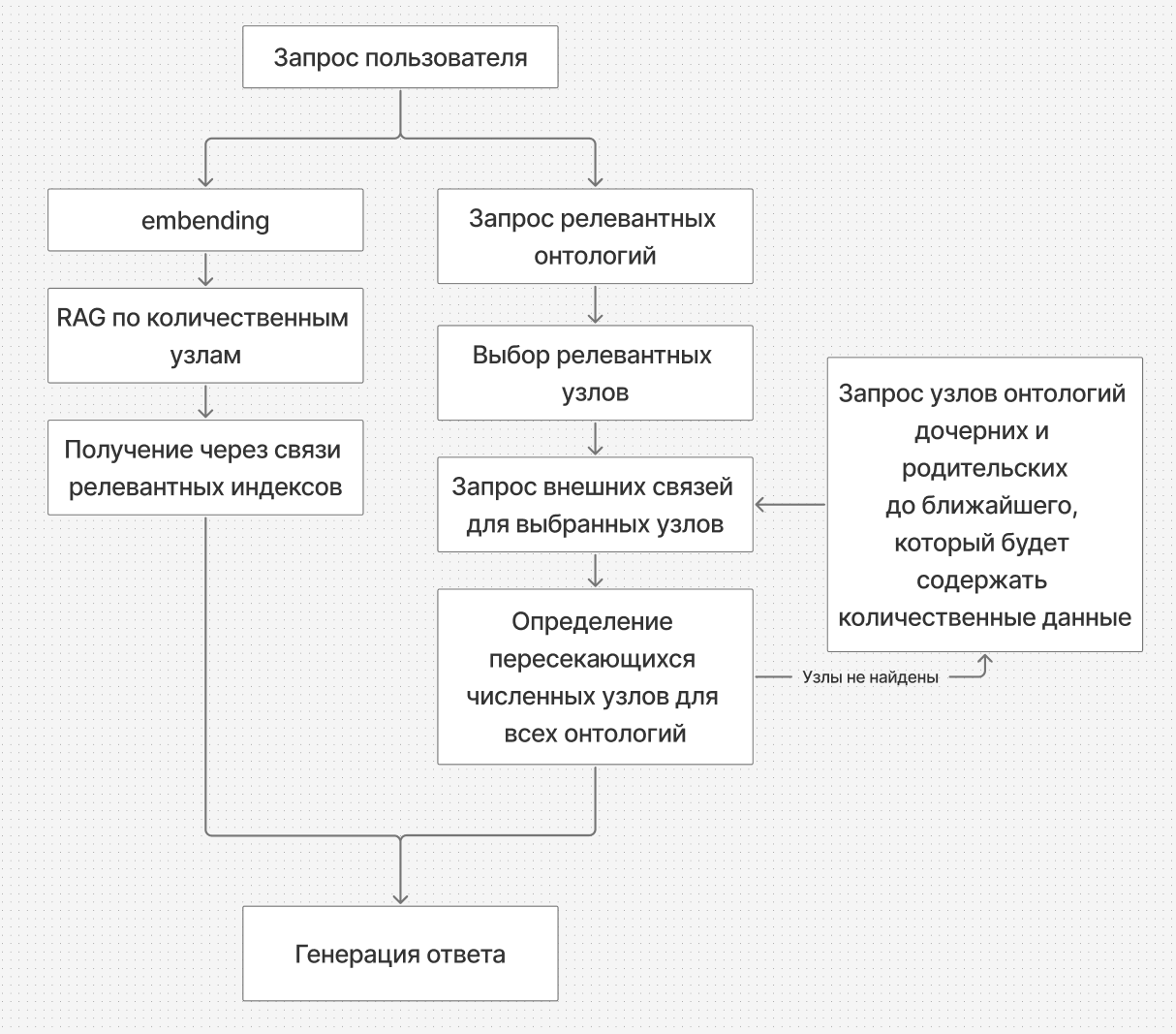
**Построение полного индекса**



В аналитическом смысле прямые количественные данные редко имеют самостоятельную ценность. Например, данные о количестве классов и численности преподавателей сами по себе не предоставляют возможность качественной интерпретации. Однако, при сопоставлении этих сведений с данными о количестве обучающихся, возможно получить информацию о среднем количестве учеников в классе и среднем количестве учеников на одного учителя. Использую данные НПА и педагогических исследований возможно методологически установить граничные значения, которые позволят интерпретировать количественные данные. Например, для среднего количества учеников в классе:

* n <15 – низкая бюджетная эффективность
* 5<n<25 – нормальное значение
* 25<n – избыточное количество учеников, снижается эффективность образовательного процесса

**Поиск количественных данных**



Метод поиска релевантных данных в БД использует:

* RAG для сопоставления эмбендингов запроса пользователя и формулировок количественных показателей
* Формирование контекста из связей между узлами

Источники информации указаны в метаданных узлов и для идентичных данных будут отображаться в контексте llm в соответствии с запрошенными онтологиями.

**Интеллектуальная визуализация данных**

Дашбоард ограничен перечнем шаблонов графиков возможных к генерации и содержит правила их использования. При проектировке схемы БД учитываются предполагаемые аналитические срезы (например, распределение показателя по регионам для федеральной карты). При таком подходе для llm построение дашбоарда сводится к выбору релевантных аналитических срезов из доступных (по наличию данных в БД). Дальнейшее построение и извлечение количественных данных происходит средствами алгоритмической обработки информации.

**Исключение галлюцинаций в работе llm**

*Пример (нетривиальны). В БД находится информация только о студентах, но не содержит информацию о школах, СПО и ДПО. На запрос пользователя о численности обучающихся llm использует данные из БД для ответа, не уточняя отсутствующие данные.*

Наличие полных онтологий в соответствии с НПА (например, иерархия и связи уровней образования) позволяет llm не только предоставить данные из БД в точной формулировке, но и определить недостающие данные в соответствии с запросом пользователя.

**Анализ и рекомендации**

Оценка предлагаемых проектов. Из текста проекта средствами NLP выделяются количественные показатели. Определяются количественные показатели наиболее близкие к выделенным целевым показателям проекта. На основе предложенных количественных изменений пересчитывается полный индекс для измерения влияния проекта на сферу образования.

Сценарии "что-если". Из предлагаемого сценария определяются количественные показатели подверженные изменению. Пересчитывается полный индекс для измерения влияния сценария на сферу образования.

Перечень возможных решений. Формируется из статичного справочника возможных решений созданного на основе анализа программ развития образования.

Оценка размера рынка образования и его сегментов.

Конкурентный анализ. Для описанного продукта из списка образовательных организаций, формируемого из анализа неструктурированных данных сайтов образовательных организаций, подбираются релевантные примеры по векторному сравнения эмбендингов запроса и данных БД.

## Наименование и производитель планируемой к замещению продукции (для проектов, направленных на импортозамещение).

https://www.holoniq.com/

## Имеющийся у коллектива участника отбора научно-технический задел по тематике заявляемого проекта (результаты исследований и испытаний прототипа/экспериментального образца продукта, техническая документация на прототип/экспериментальный образец продукта и иное), полученные ранее результаты (в т.ч. указываются документы, подтверждающие право участника отбора на интеллектуальную собственность по тематике проекта).

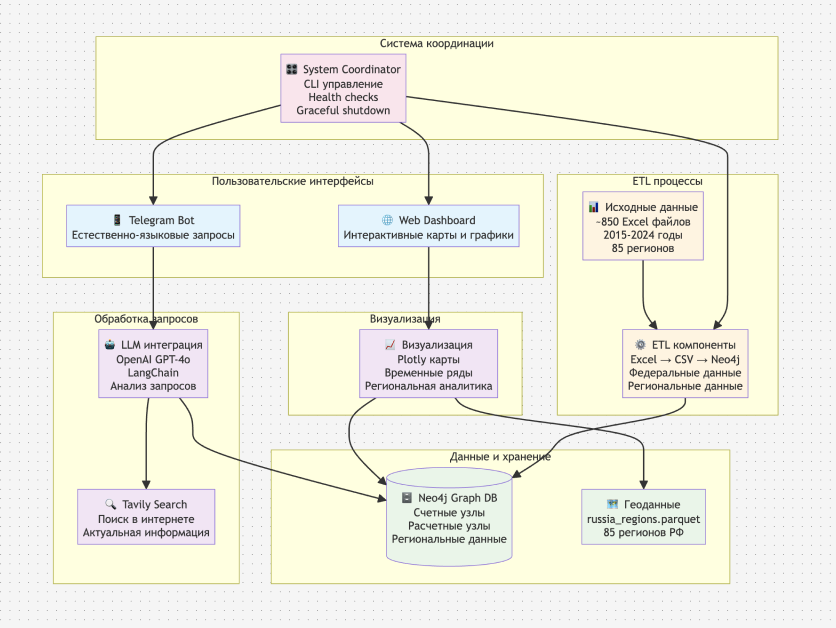
**Прототип 1 (общий) – основные функции**

Система анализа статистических форм представляет собой комплексную платформу для интеллектуального анализа образовательных данных Российской Федерации за период 2015-2024 годы. Прототип обеспечивает автоматизированную обработку статистических форм ОО-1 по всем 85 субъектам РФ, включая федеральные и региональные данные общим объемом около 10 лет статистической отчетности.

Ключевой особенностью системы является интеграция современных технологий искусственного интеллекта с традиционными методами обработки данных. Telegram-бот с поддержкой больших языковых моделей (GPT-4o) позволяет пользователям делать естественно-языковые запросы к статистическим данным, получая интеллектуальные ответы с автоматическим поиском релевантной информации через Neo4j графовую базу данных или внешние источники. ETL-компонент обрабатывает сложные Excel-файлы с объединенными ячейками и многоуровневыми структурами, создавая узлы в графовой базе данных для последующего анализа.

Веб-интерфейс системы предоставляет интерактивные карты России с региональными показателями и временные ряды данных через Flask Dashboard. Все компоненты координируются центральным System Coordinator. Архитектура построена на модульном принципе с поддержкой как локального, так и облачного развертывания через Neo4j Aura Cloud.

<https://github.com/KochemirovSergey/stat_forms>



|  |  |
| --- | --- |
| Пример запроса данных для данных находящихся в БД    Пример запроса данных для данных отсутствующих в БД |  |

**Прототип 2 – агент-аналитик**

Новый прототип системы — это рефакторинг старой версии на базе фреймворка LangGraph. Теперь ключевые методы и обработчики из прежних монолитных модулей оформлены как отдельные инструменты (“tools”), которые интеллектуальный агент может вызывать по мере необходимости — например, для запроса онтологий, данных или генерации диаграмм. Введено общее типизированное состояние (OntologyState), позволяющее передавать результаты работы и вспомогательную информацию между узлами, что повышает прозрачность и управляемость всей логики.

Что уже перенесено:

* Логика интеллектуального анализа и маршрутизации запросов пользователя (классификация запроса как количественного, образовательного или агентского).
* Доступ к базе Neo4j и инструменты для получения структуры онтологий и конкретных данных, генерация mermaid-диаграмм на лету.
* Работа с внешним поиском данных (например, Tavily).
* Централизованный контроль итераций и финальной агрегации результата в узле-оценщике.

Что еще не перенесено:

* Интерактивные дашборды и визуализация на Flask/Plotly остаются в старой архитектуре и пока отсутствуют в LangGraph-прототипе.
* ETL-модули (парсинг и загрузка Excel) не интегрированы: в будущем этот слой будет выделен в отдельного агента-парсера, взаимодействующего с основным агентом.
* Telegram-бот и веб-интерфейс нового типа не реализованы.
* Нет региональной детализации и работы со всеми 85 субъектами РФ в этом прототипе.

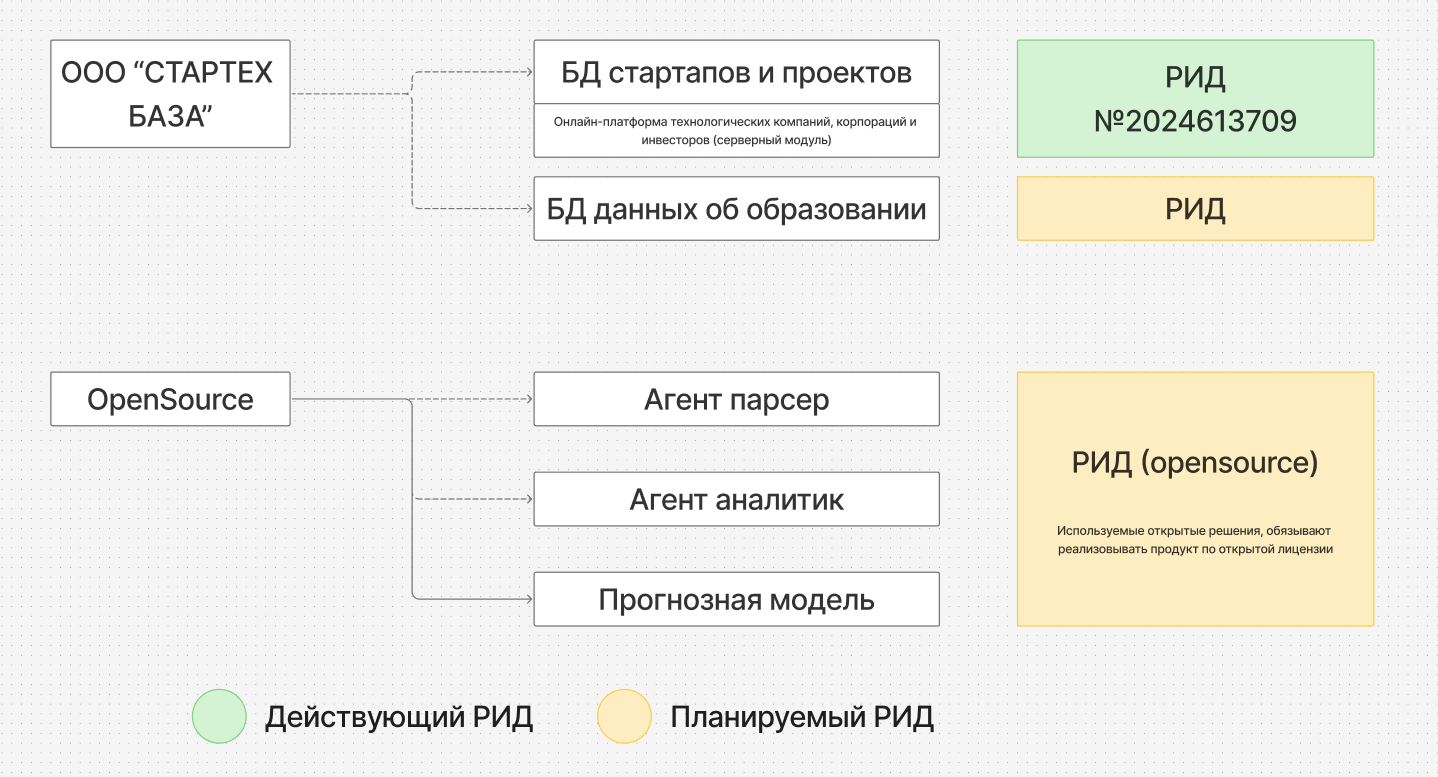
Таким образом, новый прототип — это модульное “ядро” для интеллектуальной логики и работы с онтологиями, с переносом инструментальных методов из старого монолита и закладкой основной архитектурной шины для интеграции будущих компонентов (ETL, дашборды, интерфейсы) в агентном формате.

<https://github.com/KochemirovSergey/agent_analitics>

**РИД**

## Планы по созданию и защите интеллектуальной собственности.

**Авторские права**



# Перспективы коммерциализации

## Объем и емкость рынка продукта, анализ современного состояния и перспектив развития отрасли, в которой реализуется проект (*если рынок новый, необходимо представить его описание*).

**Рынок EdTech, сегмент Market Intelligence**

|  |  |  |
| --- | --- | --- |
| Среднее значение размера мирового рынка | 218,31 | млрд долларов |
| Среднее значение CAGR мирового рынка | 14,49 | % |
| Среднее значение рынка РФ | 145 | млрд рублей |
| Среднее значение CAGR рынка РФ | 21,5 | % |
| Среднее значение размера мирового рынка, сегмент EdTech — Market Intelligence | 15 | млрд долларов |
| Среднее значение размера рынка РФ, сегмент EdTech — Market Intelligence | 8,33 | млрд рублей |
| Доля сегмента на мировом рынке | 10 | % |
| Доля сегмента на рынке РФ | 5,74 | % |
| Доля ранка РФ в мировом рынке | 0,83 | % |

**Рынок GovTech Сегмент: AI**

|  |  |  |
| --- | --- | --- |
| Среднее значение размера мирового рынка | 344 | млрд долларов |
| Среднее значение CAGR мирового рынка | 12 | % |
| Среднее значение рынка РФ | 1361,5 | млрд рублей |
| Среднее значение CAGR рынка РФ | 17 | % |
| Среднее значение размера мирового рынка, GovTech — Сегмент "AI" (AI in Government/Public Services) | 14,57 | млрд долларов |
| Среднее значение размера рынка РФ, GovTech — Сегмент "AI" (AI in Government/Public Services) | 48,66 | млрд рублей |
| Доля сегмента на мировом рынке | 4,23 | % |
| Доля сегмента на рынке РФ | 3,57 | % |
| Доля ранка РФ в мировом рынке | 4,95 | % |
| Доля сегмента РФ в моровом сегменте | 4,17 | % |

**Рынок Science AI**

|  |  |  |
| --- | --- | --- |
| Среднее значение размера мирового рынка | 6,38 | млрд долларов |
| Среднее значение CAGR мирового рынка | 20 | % |
| Среднее значение рынка РФ | 35 | млрд рублей |
| Среднее значение CAGR рынка РФ | 22,62 | % |
| Доля ранка РФ в мировом рынке | 6,86 | % |

**Доля ВВП Российской Федерации в мировой экономике**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Год** | **Показатель** | **Доля РФ в мировом ВВП** | **Описание** | **Ссылка на источник** |
| 2024 | ВВП по ППС | 3,55% | Четвёртое место в мире, данные МВФ | [wikipedia+2](https://ru.wikipedia.org/wiki/%D0%AD%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B8%D0%BA%D0%B0_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8) |
| 2024 | ВВП по ППС | 3,52% | Оценка по Statista | [statista](https://www.statista.com/statistics/271379/russias-share-of-global-gross-domestic-product-gdp/) |
| 2023 | ВВП по ППС | 2,9% | Ранее оценка МВФ | [rbc+2](https://www.rbc.ru/economics/22/10/2024/6717ac329a79478792f175ec) |
| 2025 | ВВП по ППС (прогноз/оценка) | 3,5% | Оценка World Economics | [worldeconomics](https://www.worldeconomics.com/Share-of-Global-GDP/Russia.aspx) |
| 2023 | ВВП номинальный | 1,97% | Официальная статистика Всемирного банка | [theglobaleconomy+1](https://www.theglobaleconomy.com/Russia/gdp_share/) |
| 2024 | ВВП номинальный | 2,05% | Оценка TradingEconomics | [tradingeconomics](https://tradingeconomics.com/russia/gdp) |
| 2023 | ВВП номинальный | 2,03% | Worldometers | [worldometers](https://www.worldometers.info/gdp/gdp-by-country/) |

**Итоговая оценка рынка в РФ для продукта**

Оценки мирового рынка являются более точными. Предполагаем что оценки для рынков и сегментов превышающих 2% являются завышенными и ограничим их 2%. Полученные значения размеров рынка считаем пересекающимися в оценке и в качестве искомого значения будем использовать среднее значение, а не сумму.

Итого получаем: EdTech Market Intelligence – 8,3 млрд. рублей, GovTech AI – 23,3 млрд, рублей, Science AI – 10,2 млрд. рублей.

**Средне значение оценки размера рынка в РФ для продукта – 13,9 млрд. рублей.**

Расчеты и источники информации приведены в таблице приложении.

Оценка рынка снизу от конкурентов

## Сравнение технико-экономических характеристик (включая количественные, качественные и стоимостные характеристики продукции) создаваемого инновационного продукта с зарубежными и отечественными аналогами (Таблица № 2):

**Аналоги по бизнес-модели:**

Аналог 1 - Консультант+ справочно-правовая система

<https://www.consultant.ru/>

Аналог 2 - Руспрофайл бизнес-аналитическая платформа для поиска и анализа информации о компаниях РФ

<https://www.rusprofile.ru/>

**Аналоги по используемым технологиям:**

Аналог 3 – INVENTORUS интеллектуальная аналитическая платформа для ускорения и оптимизации процессов поиска и анализа мировой научно-технической информации

<https://inventorus.ru/>

<https://platform.inventorus.ru/>

**Аналоги по обрабатываемой информации:**

Аналог 4 - Статистический сборник ежегодной серии «Индикаторы образования». Информационно-аналитическая система «Индикаторы образования».

<https://issekdash.hse.ru/viewer/public?dashboardGuid=f1b4484eecee40699e1036e6033125cb>

<https://www.hse.ru/primarydata/io>

Аналог 5 – Официальная статистика Федеральной службы государственной статистики

<https://rosstat.gov.ru/folder/10705>

<https://bi.gks.ru/biportal/contourbi.jsp?allsol=1&solution=Dashboard&project=%2FDashboard%2FStatistics_of_education_science_innovation_copy>

Таблица № 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Технико-экономические характеристики продукта** | **Аналог 1** | **Аналог 2** | **Аналог 3** | **Аналог …** | **Создаваемый продукт** |
| 1. … |  |  |  |  |  |
| 1. … |  |  |  |  |  |
| … |  |  |  |  |  |
| n. *Стоимость продукта/услуги* |  |  |  |  |  |

Дашборад ( в таблице просто)

Разбиение по статьям

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Технико-экономические характеристики продукта** | **Аналог 1** | **Аналог 2** | **Аналог 3** | **Аналог 4** | **Аналог 5** | **Создаваемый продукт** |
| 1. Количество обрабатываемых данных в целом | 348 млн., НПА | 39 источников информации:  ФНС, Росстат, Верховный суд, Казначейство, Профильные лицензирующие ведомства | Патенты:  170 млн.  Публикации:  267 млн  Авторы:  114 млн  Организаций:  12 млн | Статистические формы: |  | Все статистические формы |
| 1. Количество обрабатываемых данных по образованию | Нет информации | Реестр лицензий Рособрнадзора | Нет специальной коллекции | Количество дашбоардов: 57 | Количество дашбоардов:  8 | Динамическое построение |
| 1. Степень использование технологий ИИ | Нет информации | Система рекомендаций | Агент ассистент и генерация отчетов (RAG) | Нет | Нет | Максимальная точность количественных данных с интеллектуальной аналитикой (GraphRAG для учета статистических данных) |
| 1. Целевая аудитория | Юристы, **чиновники** | Закупщики, тендеры, **гранты** | Ученые, наукоемкие предприятия,  **гранты** | Ученые, **чиновники** | Ученые, **чиновники** | **Чиновники**, ученые, продуковенры образовательных продуктов |
| … |  |  |  |  |  |  |
| n. *Стоимость продукта/услуги* | 1000 рублей в месяц для физ лиц  10 000 рублей в месяц для юридических лиц | 470 рублей в месяц | 3 750 рублей в месяц | Бесплатно | Бесплатно | Для граждан – бесплатно  Для нии и образовательных организаций– 3000 в месяц  Для органов управления – 10-15 млн. в год |

Целевые сегменты потребителей создаваемого продукта и оценка платежеспособного спроса. Потенциальные клиенты (заказчики).

* 1. Гражданские запросы, обращения граждан – платежеспособность 0, но на общественных принципах считаем важным полностью удовлетворять потребность. Берем в учет оценку (низкая) количества необходимых запросов на одного пользователя с customer joyrney map «Гражданские» и в замен просим отвечать на опросную анкету – обогащаем базу данных, повышаем ценность. Опросник делаем большой (может даже генерируемый с повышением сложности). В убыток, но убеждаем в необходимости покупки.
  2. Вузы, ДПО, другие участники рынка – 3-5 тыс. в месяц, если сравнивать с руспрофайлом и консультантом. Эта БД все-таки более специфичная и реже применимая

Список

* 1. Органы управления – 15-20 млн. в год, сам продукт бесплатно + за локальную обработку внутренней документации.

**Планируемая стоимость продукта. Расчет себестоимости.**

**Серверные издержки статичные**

**Издержки api**

**Вызовов llm на итерацию = 2**

**Итераций = 3**

**Оценка точки окупаемости для тарифа**

**Издрежки api на обработку документов**

## Описание бизнес-модели проекта . Производственный план и план продаж.

|  |  |
| --- | --- |
| **Блок** | **Описание** |
| **1. Потребительские сегменты (Customer Segments)** | **Граждане** — бесплатное использование, источник данных и обратной связи. **Малые образовательные и научные организации** (НИИ, вузы, ДПО) — низкая подписка, ориентир на массовый спрос. **Органы управления / крупные организации** — высокая ценность аналитики, интеграция с внутренними системами, основная прибыль. |
| **2. Ценностные предложения (Value Proposition)** | **Граждане:** доступ к аналитике, участие в развитии проекта через анкетирование. **Малые организации:** доступная цена, уникальные аналитические и прогнозные инструменты, конкурентные преимущества. **Крупные клиенты/органы управления:** комплексный анализ открытых и внутренних данных, прогнозирование, сценарное моделирование, интеграция с внутренними системами, приватность, консалтинг. |
| **3. Каналы сбыта (Channels)** | **Академические каналы**: публикации и реклама в научных журналах, участие в конференциях и профильных мероприятиях.  **Прямые контакты**: теплые обзвоны и рассылки по наработанной базе образовательных учреждений и органов управления.  **Цифровые каналы**: Телеграм-бот и веб-портал с дашбордами, интеграция в корпоративные системы. |
| **4. Взаимоотношения с клиентами (Customer Relationships)** | **Граждане:** самообслуживание, геймификация через опросы, вовлечение в развитие. **Малые организации:** техническая поддержка, обучение, консультации. **Крупные клиенты:** персональное сопровождение, кастомизированные решения, закрытые презентации. |
| **5. Потоки доходов (Revenue Streams)** | Бесплатная версия для граждан (косвенная ценность — база данных и вовлечение).  Подписка для малых НИИ/вузов/ДПО по сниженной цене (массовый спрос).  "Про" пакеты и индивидуальные проекты для органов управления — обработка внутренних документов, прогнозирование, аналитика. |
| **6. Ключевые ресурсы (Key Resources)** | Собственная база данных, алгоритмы анализа неструктурированных данных, графовая база и индексы, команда разработчиков и аналитиков, инфраструктура для быстрого ответа, сеть теплых контактов. |
| **7. Ключевые виды деятельности (Key Activities)** | Анализ и интеграция данных, построение прогнозных моделей и индексов, разработка и поддержка ПО (бот, веб), поддержка клиентов и обучение, кастомизация под крупные проекты. |
| **8. Ключевые партнеры (Key Partners)** | Провайдеры данных, академические и отраслевые организации, технологические платформы для обработки данных, эксперты в области аналитики и статистики. |
| **9. Структура издержек (Cost Structure)** | Разработка и поддержка платформы, автоматизация загрузки данных, кастомизация под крупных клиентов, затраты на публикации и участие в конференциях, маркетинг в научной среде, техническая поддержка. |