

Курсовая работа на тему «Разработка автоматической диалоговой системы на основе языковой модели в сфере бизнес-аналитики»

Выполнила Кызыл-оол Монгун-Ай ПМ22-6

Научный руководитель: Петрунина Е.В.

Введение

Бизнес-аналитика — сфера деятельности на стыке экономики, ИТ и менеджмента, которая занимается процессом систематического анализа данных для принятия управленческих решений с целью снижения рисков и выявления новых возможностей для бизнеса.

Бизнес-аналитики занимаются анализом данных, выявляя тенденции и закономерности для улучшения бизнес-процессов. Они работают с финансовыми показателями, прогнозируют доходы и расходы, оптимизируют операции, автоматизируют отчётность и выявляют узкие места. Они визуализируют данные в дашбордах, готовят аналитические отчёты для руководства. Также анализируют документы, выявляют риски и строят прогнозные модели.

Цель проекта:

создание интеллектуального интерфейса, который позволит пользователю, не обладая навыками программирования или технического анализа, оперативно взаимодействовать с бизнес-данными и получать осмысленные ответы на свои вопросы

Актуальность

В рамках данной курсовой работы реализуется интерактивная интеллектуальная система, способная вести диалог с пользователем, интерпретировать его запросы, анализировать документы и строить визуальные отчёты на лету.

Это особенно актуально для компаний, где важно автоматизировать обработку отчётов, ускорить аналитику и снизить зависимость от специалистов по данным.

Объект исследования: процесс автоматизации аналитической обработки бизнес-информации с использованием технологий искусственного интеллекта, в частности — языковых моделей

Предмет исследования: методы и архитектурные решения, применяемые при создании диалоговой системы, способной понимать естественный язык

Задачи:

1. Обработка естественного языка
2. Определение намерения пользователя
3. Извлечение данных из текстов и PDF-документов
4. Приведение данных к табличной структуре
5. Визуализация данных
6. Смысловой поиск по базе документов
7. Разработка диалоговой системы

Данные:

1. Текстовый отчет
2. Выжимки из статей по бизнес-аналитике
3. PDF файл с отчетом

Описание моделей

Обе модели — часть открытой линейки **Gemma** от Google, ориентированной на приватность и использование вне облака.

Gemma 3-12B IT-QAT

- **Параметров:** ~12 млрд
- **Тип:** Инструкция-файнтюнинг (Instruction-tuned) + калиброванная с помощью QAT (Quantization Aware Training)
- **Качество:** Высокая точность, уверенно работает с аналитическими задачами, PDF-документами и числовыми данными
- **Требования:** Требуется GPU или мощный сервер для запуска
- **Применение:** Подходит для глубокой аналитики, генерации текстов, ответов на сложные вопросы и задач со структурированными данными

Gemma 2B

- **Параметров:** ~2 млрд
- **Тип:** Компактная LLM
- **Качество:** Средняя точность, справляется с простыми задачами и короткими текстами
- **Требования:** Запускается на CPU, подходит для локального и мобильного использования
- **Применение:** Быстрые ответы, базовая аналитика, разработка на слабых устройствах или при ограниченных ресурсах

Определение намерений пользователя

эта функция анализирует сообщение пользователя и определяет:

- нужно ли строить график (< CALL >)
- нужно ли искать информацию в базе данных (< SEARCH >)

для вывода ответа используется `ask_gemma_simple`, которому передается системный промпт

```
[ ] def get_intent(message):
    sys_prompt="""Ты – помощник, который должен определить, что хочет пользователь, по его сообщению.
    У тебя есть два параметра, которые нужно установить:\n\n
    <CALL> – если пользователь явно просит построить график, установи <CALL> = 1, иначе <CALL> = 0.\n
    <SEARCH> – если пользователь прямо указывает, что нужно что-то найти в базе данных
    (например: 'найди', 'проверь в базе', 'покажи информацию о...'), установи <SEARCH> = 1, иначе <SEARCH>
    Ты должен вернуть ответ только в следующем формате:\n<CALL> = [0 или 1]\n<SEARCH> = [0 или 1]\n\n

    Примеры:\n- Пользователь: 'Построй график продаж за месяц' → <CALL> = 1, <SEARCH> = 0\n
    - Пользователь: 'Найди сотрудника по фамилии Петров' → <CALL> = 0, <SEARCH> = 1\n
    - Пользователь: 'Покажи информацию о клиенте Иванов' → <CALL> = 0, <SEARCH> = 1\n
    - Пользователь: 'Сколько было заказов в марте?' → <CALL> = 0, <SEARCH> = 0\n\n

    Теперь проанализируй следующее сообщение пользователя и верни ответ в указанном формате:"""
```

```
res=ask_gemma_simple(message,sys_prompt=sys_prompt,max_tokens=20)
return res
```

```
[ ] print(get_intent('Привет найди в базе информацию о бизнес аналитике'))
print('-'*100)
print(get_intent('Привет построй график прибыли и налогов'))
print('-'*100)
print(get_intent('Привет как дела?'))
```

```
☞ <CALL> = 0
   <SEARCH> = 1
   -----
   <CALL> = 1
   <SEARCH> = 0
```

```
-----
<CALL> = 0
<SEARCH> = 0
```

Варианты намерений:

1. <CALL> = 1 — построение графика;
2. <SEARCH> = 1 — выполнение поиска по базе;
3. <CALL> = 0, <SEARCH> = 0 — общие информационные запросы (например, «что такое EBITDA?»).

Работа с текстовым отчетом

```
test_p'''Финансовый отчет компании "Альфа Тех" за 2019-2023 гг. показывает устойчивый рост.
```

```
Выручка компании увеличивалась ежегодно:
```

```
в 2019 году она составила 125 млн рублей,  
в 2020 – 138 млн,  
в 2021 – 154 млн,  
в 2022 – 167 млн,  
в 2023 – 180 млн рублей.
```

```
Расходы за тот же период:
```

```
2019 – 92 млн,  
2020 – 100 млн,  
2021 – 113 млн,  
2022 – 120 млн,  
2023 – 129 млн рублей.
```

```
Чистая прибыль по годам:
```

```
2019 – 33 млн,  
2020 – 38 млн,  
2021 – 41 млн,  
2022 – 47 млн,  
2023 – 51 млн рублей.
```

```
Капитализация на конец 2023 года достигла 750 млн рублей. В компании отмечают рост маржи и снижение долговой нагрузки.'''
```

На его основе формируем JSON для дальнейшей обработки

Вспомогательные функции:

``clean_and_parse_json_block`` - убирает лишние символы, парсится строка и возвращается валидный JSON внутри `<json >...< /json>`.

``extract_and_send_json_for_fix(json_text)`` — исправление "кривого" JSON

Основная функция для контекстного ответа с извлечением данных:

- Принимает вопрос пользователя и контекст (пример - финансовый отчет)
- Анализирует текст на наличие числовых данных
- Формирует ответ с извлеченными данными в формате JSON
- Возвращает текстовый ответ и DataFrame с данными

```
def ask_with_context(message, context=None):
    import re
    import pandas as pd

    need_df = 0
    err = 0

    messages = []

    messages.append({
        "role": "system",
        "content": (
            "Ты — ассистент по бизнес-аналитике. Отвечай строго по предоставленному тексту, "
            "если он есть. Если в нём есть числовые или табличные данные (например, финансовая отчётность), "
            "в конце добавь один валидный блок <json>...</json> с таблицей, которую можно загрузить в pandas.DataFrame.\n\n"

            "Если текста нет или он не нужен, ты можешь отвечать на общеизвестные вопросы самостоятельно.\n\n"

            "Правила для JSON-блока:\n"
            "1. Используй только один блок <json>...</json>.\n"
            "2. Структура — список словарей (каждый словарь — строка таблицы).\n"
            "3. Все значения должны быть числовыми (int или float), кроме заголовков показателей.\n"
            "4. Не добавляй пояснений внутри блока.\n\n"

            "Пример:\n"
            "<json>\n"
            "{\n"
            "  \"Баланс\": [\n"
            "    {\"Показатель\": \"Активы\", \"2022\": 282418, \"2023\": 320728},\n"
            "    {\"Показатель\": \"Внеоборотные активы\", \"2022\": 249701, \"2023\": 270065}\n"
            "  ]\n"
            "}\n"
            "</json>"
        )
    })

    if context:
        messages.append({
            "role": "user",
            "content": f"Вот часть текста документа:\n{context}"
        })

    messages.append({
        "role": "system",
        "content": "Если документа нет, ты можешь использовать общие знания и дать корректный ответ."
    })

    messages.append({
        "role": "user",
        "content": f"Вопрос: {message}"
    })
```

Алгоритм работы функции:

Формирование prompt с системой и пользователем:

- если есть контекст — вставляем текст документа;
- добавляем вопрос;
- обучаем модель возвращать JSON в `<json>...</json>` блоке.

Отправляем запросы messages в модель Gemma (через `ask_gemma_simple()`).

Парсим JSON из ответа:

- Ищем `<json>...</json>`;
- Чистим строку (через `extract_and_send_json_for_fix`);
- Преобразуем в dict (`clean_and_parse_json_block`);
- Создаем `pandas.DataFrame` из вложенного словаря.

Возвращает результат:

- Если таблица есть — (текст_до_json, df);
- Если таблицы нет — просто строку-ответ.

```
def ask_with_context(message, context=None):
    import re
    import pandas as pd

    need_df = 0
    err = 0

    messages = []

    messages.append({
        "role": "system",
        "content": (
            "Ты — ассистент по бизнес-аналитике. Отвечай строго по предоставленному тексту, "
            "если он есть. Если в нём есть числовые или табличные данные (например, финансовая отчётность), "
            "в конце добавь один валидный блок <json>...</json> с таблицей, которую можно загрузить в pandas.DataFrame.\n\n"

            "Если текста нет или он не нужен, ты можешь отвечать на общеизвестные вопросы самостоятельно.\n\n"

            "Правила для JSON-блока:\n"
            "1. Используй только один блок <json>...</json>.\n"
            "2. Структура — список словарей (каждый словарь — строка таблицы).\n"
            "3. Все значения должны быть числовыми (int или float), кроме заголовков показателей.\n"
            "4. Не добавляй пояснений внутри блока.\n\n"

            "Пример:\n"
            "<json>\n"
            "{\n"
            "  \"Баланс\": [\n"
            "    {\"Показатель\": \"Активы\", \"2022\": 282418, \"2023\": 320728},\n"
            "    {\"Показатель\": \"Внеоборотные активы\", \"2022\": 249701, \"2023\": 270065}\n"
            "  ]\n"
            "}\n"
            "</json>"
        )
    })

    if context:
        messages.append({
            "role": "user",
            "content": f"Вот часть текста документа:\n{context}"
        })

    messages.append({
        "role": "system",
        "content": "Если документа нет, ты можешь использовать общие знания и дать корректный ответ."
    })

    messages.append({
        "role": "user",
        "content": f"Вопрос: {message}"
    })
```

Построение графика:

Эта функция:

- Принимает DataFrame и вопрос пользователя
- Определяет, какие колонки нужны для ответа на вопрос
- Строит соответствующий график с помощью matplotlib
- Возвращает список используемых колонок и фигуру с графиком

```
def call(df, question):
    columns = list(df.columns)
    table_text = df.to_string(index=False)

    messages = [
        {
            "role": "system",
            "content": ""Ты — аналитик данных. Твоя задача — по таблице и вопросу пользователя определить,
            какие именно колонки из таблицы нужны для ответа.

Ты должен:
1. Проанализировать таблицу и вопрос.
2. Выбрать только те колонки, которые действительно необходимы.
3. Вернуть их в следующем формате (без пояснений и комментариев): < ['col_name', 'col_name', ...] >

Пример:
Таблица:
    Год  Выручка  Расходы  Чистая прибыль
0 2019  125000000  92000000   33000000
1 2020  138000000  100000000   38000000
2 2021  154000000  113000000   41000000
3 2022  167000000  120000000   47000000
4 2023  180000000  129000000   51000000

Доступные колонки: ['Год', 'Выручка', 'Расходы', 'Чистая прибыль']

Вопрос пользователя:
"построй график выручки и прибыли"

Ответ:
['Выручка', 'Чистая прибыль']
"""

        },
        {
            "role": "user",
            "content": f""Вот таблица данных:

{table_text}

Вот доступные колонки: {columns}

Вопрос пользователя:
\"{question}\"
"""

    }
```

Результат работы

```
answer=ask_with_context('Сделай выводы о прибыли и расходах',test_p)
```

```
answer[0]
```

'Согласно предоставленному финансовому отчету, компания "Альфа Тех" демонстрирует устойчивый рост как выручки, так и расходов в период с 2019 по 2023 год. Чистая прибыль также последовательно увеличивалась за этот период.\n\n* **Выручка:** Ежегодный прирост выручки от 125 млн рублей в 2019 году до 180 млн рублей в 2023 году.\n\n* **Расходы:** Расходы также росли, но более медленными темпами по сравнению с выручкой, что указывает на улучшение эффективности управления затратами.\n\n* **Чистая прибыль:** Чистая прибыль последовательно увеличивалась от 33 млн рублей в 2019 году до 51 млн рублей в 2023 году, что свидетельствует о повышении прибыльности компании.\n\n'

```
df = answer[1]
```

```
answer[1].columns
```

```
Index(['Год', 'Выручка', 'Расходы', 'Чистая прибыль'], dtype='object')
```

```
display(df)
```

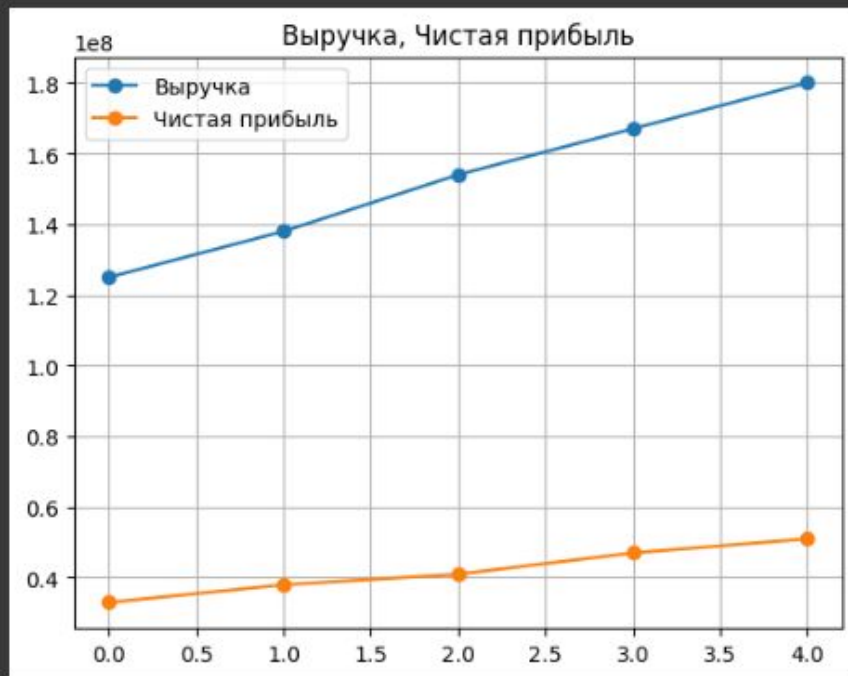
	Год	Выручка	Расходы	Прибыль
0	2019	125	92	33
1	2020	138	100	38
2	2021	150	110	41
3	2022	165	120	47
4	2023	180	130	51

```
answer, fig = call(df, "построй график выручки и чистой прибыли")
```

```
answer
```

```
"['Выручка', 'Чистая прибыль']"
```

```
fig
```



Работа с выжимками статей

С помощью пакета
langchain_community.document_loaders
извлекаем тексты статей из Википедии и
популярных сайтов по бизнес-аналитике (через
WikipediaLoader и WebBaseLoader)

Всего извлечено 198 фрагментов.

```
urls = [  
    "https://www.forbes.ru/tegi/biznes-analitika",  
    "https://www.it-world.ru/tag/bi/",  
    "https://www.batimes.com/",  
    "https://business-analytics-russia.ru/articles-for-business-analyst/",  
    "https://www.dbta.com/Categories/Business-Intelligence-and-Analytics-327.aspx"  
]
```

```
all_chunks = wiki_chunks + web_chunks  
print(f"Всего фрагментов: {len(all_chunks)}")  
print(all_chunks[0].page_content[:1000]) # вывод примера содержимого
```

Всего фрагментов: 198

Бизнес-аналитик – специалист, использующий методы бизнес-анализа для исследования потребностей деятельности. Международный Институт Бизнес-Анализа (IIBA, International Institute of Business Analysis) определяет бизнес-аналитика в консалтинговом бизнесе бизнес-аналитиком называется высшая позиция для консультанта. IIBA отмечает всемирный день бизнес-анализа 1-го ноября

== См. также ==

Аналитическая записка
SWOT-анализ
PEST-анализ

Извлечение выжимок

```
prompt_template = """Напиши краткое содержание следующего текста на русском языке:

{text}

КРАТКОЕ СОДЕРЖАНИЕ НА РУССКОМ:"""
PROMPT = ChatPromptTemplate.from_template(prompt_template)

summary = summarize_chunks(all_chunks[:10])
print(summary)

Бизнес-аналитик – специалист, обеспечивающий сбор, анализ, коммуникацию и проверку

with open("summaries_ru.txt", "w", encoding="utf-8") as f:
    for i, doc in enumerate(all_chunks):
        f.write(f"--- Document {i} ---\n")
        summary = summarize_chunks([doc])
        f.write(summary + "\n\n")
```

Просим модель написать краткое содержание статей на русском и английском, сохраняем результат в текстовые файлы.

```
{'Document 0': "Бизнес-аналитик - специалист, применяющий методы бизнес-анализа для исследования потребностей деятельности организаций с целью определения проблем бизнеса и предложений их решения. Международный Институт Бизнес-Анализа (ИИБА) определяет бизнес-аналитика как посредника для сбора, анализа, коммуникации и проверки требований к изменению бизнес-процессов, правил и информационных систем. Как высококвалифицированный консалтинг-специалист, бизнес-аналитик понимает проблемы и возможности бизнеса в контексте требований и рекомендует решения, позволяющие организации достичь своих целей. ИИБА отмечает всемирный день бизнес-анализа 1-го ноября. Автор использовал термины 'Аналитическая записка', 'SMOT-анализ', 'PEST-анализ', 'Анализ пяти сил Портера', чтобы обозначить различные методы бизнес-анализа.",
'Document 1': 'Аналитика - это часть искусства рассуждений или логики, затрагивающая учение о разборе целого на составляющие части при познании или предметно-практической деятельности человека.',
'Document 2': 'Аналитика в узком смысле означает систематический численный анализ данных с целью обнаружения и интерпретации значимых закономерностей. Он применяется в областях, где характерно накопление больших объёмов информации и опирается на совмещение статистики, компьютерного программирования и исследования операций. Аналитика подразделяется на описательную, диагностическую, прогнозную, предписывающую и когнитивную. Применяется в различных бизнес-сферах, включая маркетинг, управление, финансы, информационную безопасность и программное обеспечение. Для обработки больших объёмов данных алгоритмы аналитики используют методы информатики, статистики и математики.',
'Document 3': 'Текст рассказывает историческую справочную информацию о науке формальной логики, в частности, о работах Аристотеля и Иммануила Канта по этой теме и терминах, которые они использовали для своих подходов к логике и анализу познавательной способности человека. Также упоминается, что позже, чем Аристотель, термин «логика» стал установиться в обиходе.\n\nВАЖНЫЕ ТЕМЫ: формальная логика как самостоятельная наука, работы Аристотеля и Канта по логике, а также их вклад в понятие аналитики по отношению к логике и познавательной способности человека.'},
```

```
{'Document 0': "A business analyst is a specialist who investigates organizations' needs, identifies problems, and proposes solutions. They act as intermediaries between stakeholders, analyzing and communicating requirements for changing business processes and systems. The role is defined by the International Institute of Business Analysis as crucial for helping organizations achieve their goals, and in consulting, it is a senior consultant position. World Business Analysis Day is celebrated on November 1st.",
'Document 1': 'Analytica is the process of analyzing and breaking down a whole into its constituent parts through cognitive or practical human activity.',
'Document 2': 'Analytics is a methodical approach to analyzing numerical data to uncover patterns and insights, using statistics, computer programming, and operations research. It is used in various fields such as marketing, finance, and cybersecurity to make informed decisions and solve complex problems by processing large datasets with algorithms that combine computer science, statistics, and mathematics.',
'Document 3': 'The term "analytic" has a long history, dating back to Aristotle\'s use in the 4th century BCE to describe logic, and was later referenced by Immanuel Kant in the 18th century to describe the breakdown of human cognitive abilities.',
'Document 4': 'Analytics is an interdisciplinary field that uses methods such as machine learning and neural networks. It also includes unsupervised machine learning methods like cluster analysis and principal component analysis for segmentation profiling.',
'Document 5': 'Marketing involves using data analysis and customer feedback to create strategies for branding and revenue. Marketing analytics includes both qualitative and quantitative data for making strategic decisions and includes predictive modeling, experiments, automation, and real-time communication to help companies forecast and achieve maximum results.'},
```

Поиск релевантных документов по вопросу пользователя

```
def search(question, db):
    with open(db, 'r') as text:
        t=text.read()
        messages = [
            {
                "role": "system",
                "content": """Ты – аналитик данных. Твоя задача – помочь пользователю найти подходящие статьи на основе его запроса.

Тебе предоставлен документ, содержащий список файлов с кратким описанием (выжимкой) содержания каждой статьи.

Что нужно сделать:
1. Проанализируй вопрос пользователя.
2. Сравни смысл запроса с содержанием каждой статьи.
3. Верни список названий тех документов, которые наиболее соответствует запросу.
4. Если подходящих документов нет – верни пустой список: []

Формат ответа:
['<filename>', '<filename>', ...]

Дополнительно:
- Статьи и запрос могут быть на разных языках – учитывай смысл, а не язык.
- Не добавляй пояснений, комментариев или другого текста – только список файлов.
"""
            },
            {
                "role": "user",
                "content": f"""Вот база документов:

{t}

Вопрос пользователя:
"{question}"
"""
            }
        ]
    return ast.literal_eval((ask_gemma_simple(m=messages).strip()))
```

Функция ищет документы в базе , релевантные вопросу пользователя, используя языковую модель для семантического анализа.

Формирует запрос к ИИ-модели с:

- Инструкцией анализировать запрос и сопоставлять с описаниями документов
- Содержимым базы документов
- Вопросом пользователя

Возвращает список названий подходящих файлов

Результат работы

```
res = search("Найди в базе информацию об определении слова веб-аналитика", "summaries_ru.txt")
res
```

```
['Document 6']
```

```
documents = get_documents_by_ids(res)
```

```
for doc_id, content in documents.items():
    print(f'{doc_id}': {content}\n")
```

```
'Document 6': Веб-аналитика - это процесс сбора данных о действиях посетителей веб-сайта во время сеанса с целью улучшения маркетинга
```


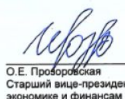
```
documents = get_documents_by_ids(res, 'summaries_eng.txt')
```

```
for doc_id, content in documents.items():
    print(f'{doc_id}': {content}\n")
```

```
Document 1: Analytica is the process of analyzing and breaking down a whole into its constituent parts through cognitive or practical
```

```
Document 4: Analytics is an interdisciplinary field that uses methods such as machine learning and neural networks. It also includes
```

Работа с PDF

ПАО НК «РуссНефть»			
Консолидированный отчет о прибыли или убытке и прочем совокупном доходе			
за год, закончившийся 31 декабря 2023 г.			
(в миллионах российских рублей)			
	Прим.	За год, закончившийся 31 декабря 2023 г.	За год, закончившийся 31 декабря 2022 г.
Выручка	10	238 725	290 812
Себестоимость реализации	11	(172 205)	(233 199)
Валовая прибыль		66 520	57 613
Расходы на геологоразведочные работы		(769)	(350)
Коммерческие расходы	12	(14 224)	(12 319)
Общехозяйственные и административные расходы	12	(3 940)	(5 038)
Прочие операционные доходы	14	3 833	7 481
Прочие операционные расходы	14	(27 081)	(13 456)
Операционная прибыль		24 439	34 911
Финансовые доходы	13	7 759	5 018
Финансовые расходы	13	(15 280)	(9 068)
Курсовые разницы, нетто		1 426	(3 246)
Прибыль до налогообложения		18 364	26 685
Расход по налогу на прибыль	17	(12 517)	(6 548)
Прибыль за отчетный период		5 847	20 137
Прочий совокупный доход, который впоследствии может быть реклассифицирован в состав прибыли или убытка			
Прибыль при пересчете иностранных валют		1 767	748
Итого совокупный доход за вычетом налогов		7 614	20 885
Прибыль(убыток), приходящийся на:			
Акционеров Материнской компании		20 441	20 446
Неконтролирующие доли участия		(14 594)	(309)
Итого совокупный доход(убыток), приходящийся на:			
Акционеров Материнской компании		21 042	21 829
Неконтролирующие доли участия		(13 428)	(644)
Прибыль на акцию – базовая и разведенная, руб.	23	33	54
Средневзвешенное количество обыкновенных акций, млн шт.		294	294
			
Е.В. Толочко			
Председатель			
Дата утверждения: 25 марта 2024 г.			
			
О.Е. Прокофьев			
Старший вице-президент по экономике и финансам			
Прилагаемые примечания являются неотъемлемой частью данной консолидированной финансовой отчетности.			

```
def extract_text_from_pdf(path: str) -> str:
    doc = fitz.open(path)
    text = ""
    for page in doc:
        text += page.get_text()
    return text
pdf_path = "test_report.pdf"
document_text = extract_text_from_pdf(pdf_path)
print(document_text[:4000])
```

ПАО НК «РуссНефть»	
Консолидированный отчет о финансовом положении	
за год, закончившийся 31 декабря 2023 г.	
(в миллионах российских рублей)	
Прилагаемые примечания являются неотъемлемой частью данной консолидированной финансовой отчетности.	
9	
Прим.	
31 декабря	
2023 г.	
31 декабря	
2022 г.	
Активы	
Внеоборотные активы	
Основные средства	

15
155 086
153 807
Активы в форме права пользования
16
721
1 003
Активы по разведке и оценке запасов
18
1
3 647
Гудвил
17
9 944
9 961

Функция анализирует бизнес-документы (в т.ч. PDF) и отвечает на вопросы пользователя, извлекая структурированные данные.

Формирует запрос к ИИ:

- Системный промпт с инструкцией для бизнес-анализа
- Требование возвращать данные в формате <json> (если есть числовая информация)

Если есть структурированные данные:

- Извлекает JSON, преобразует в pandas.DataFrame
- Возвращает текстовый ответ + таблицу

Если данных нет — возвращает только текстовый ответ

```
def ask_with_context_pdf(message, context=None):
    need_df=0
    err=0
    messages = [
        {"role": "system", "content": "Ты — ассистент по бизнес-аналитике. Ответь на вопросы пользователя, извлекая структурированные данные из PDF-документов."},
        {"role": "user", "content": f"Вот часть текста документа:\n{context}"},
        {"role": "system", "content": "Если документа нет, то отвечай не по сод"},
        {"role": "user", "content": f"Вопрос: {message}"},
    ]
    response = client.chat.completions.create(
        model="gemma:3-12b-it-qat",
        messages=messages,
        temperature=0.3,
        max_tokens=10_000
    )
    answer=response.choices[0].message.content

    need_df = '<json>' in answer
    ind=answer.find('<json>')
    if need_df:
        match = re.search(r"<json>(.*?)</json>", answer, re.DOTALL)
        err = not match
        if err:
            return 'Произошла ошибка при формировании таблицы'
        raw_json = match.group(1).strip()
        cleaned=extract_and_send_json_for_fix(raw_json)
        parsed = clean_and_parse_json_block(cleaned)
        df = pd.DataFrame(parsed[list(parsed.keys())[0]])
        return answer[:ind],df
    else:
        return answer
```

Результат работы

```
context = document_text
question = "Сделай ключевые выводы исходя из отчета"
answer = ask_with_context_pdf(question, context)

print("Ответ от Gemma:\n", answer[0])
```

Ответ от Gemma:
Согласно консолидированному отчету о финансовом положении ПАО НК «РоссНефть» за год, закончившийся 31 декабря 2023 г., можно сделать

- * **Рост активов:** Общие активы компании увеличились с 282 418 миллионов рублей на 31 декабря 2022 года до 320 728 миллионов рублей
- * **Увеличение внеоборотных активов:** Внеоборотные активы выросли с 249 781 миллиона рублей до 270 065 миллионов рублей, что обусл
- * **Рост оборотных активов:** Оборотные активы также увеличились, с 32 717 миллионов рублей до 50 663 миллионов рублей. Значительный
- * **Увеличение капитала:** Капитал, приходящийся на акционеров материнской компании, увеличился с 81 413 миллионов рублей до 91 763
- * **Рост долгосрочных обязательств:** Долгосрочные обязательства выросли с 117 260 миллионов рублей до 132 160 миллионов рублей, гл
- * **Увеличение краткосрочных обязательств:** Краткосрочные обязательства увеличились с 71 491 миллиона рублей до 93 523 миллионов ру

```
<json>
{
  "Активы": [
    {"Показатель": "Внеоборотные активы (2022)", "Сумма": 249781},
    {"Показатель": "Внеоборотные активы (2023)", "Сумма": 270065},
    {"Показатель": "Оборотные активы (2022)", "Сумма": 32717},
    {"Показатель": "Оборотные активы (2023)", "Сумма": 50663},
    {"Показатель": "Активы (2022)", "Сумма": 282418},
    {"Показатель": "Активы (2023)", "Сумма": 320728}
  ],
  "Капитал и обязательства": [
    {"Показатель": "Капитал, приходящийся на акционеров (2022)", "Сумма": 81413},
    {"Показатель": "Капитал, приходящийся на акционеров (2023)", "Сумма": 91763},
    {"Показатель": "Долгосрочные обязательства (2022)", "Сумма": 117260},
    {"Показатель": "Долгосрочные обязательства (2023)", "Сумма": 132160},
    {"Показатель": "Краткосрочные обязательства (2022)", "Сумма": 71491},
    {"Показатель": "Краткосрочные обязательства (2023)", "Сумма": 93523}
  ]
}
</json>
```

display(answer[1])

	Год	Внеоборотные активы	Оборотные активы	Активы
0	2022	249701.0	NaN	NaN
1	2023	270065.0	NaN	NaN
2	2022	NaN	32717.0	NaN
3	2023	NaN	50663.0	NaN
4	2022	NaN	NaN	282418.0
5	2023	NaN	NaN	320728.0

Разработка автоматической диалоговой системы

Вспомогательные функции:

`extract_local_pdf_links(text)` - ищет в тексте локальные ссылки на PDF-файлы. Используется, чтобы понять, не содержит ли контекст файл.

``context_handle(context)`` - если в контексте найдена ссылка на PDF, извлекает его текст, иначе возвращает исходный контекст.

parse_call_search(text) - ищет в тексте намерения - метки < CALL> /< SEARCH>, чтобы понять, нужно ли строить график (~~CALL~~=1) и/или искать статьи (SEARCH=1).

```
def extract_local_pdf_link(text):
    pattern = r'(?<!https:)(?<!http:)(?<!mailto:)(?<![/])(?<![\w:/])(?<![.~?])(?<![\w\-\./\+])?\[w\.-]+\.pdf'
    return re.findall(pattern, text, re.IGNORECASE)

def context_handle(context):
    a = extract_local_pdf_links(text)
    if len(a) == 0:
        return context
    else:
        return extract_text_from_pdf(a[0])[3500]

def parse_call_search(text):
    match = re.search(r"<CALL>\s*=\s*(\d)\s*<SEARCH>\s*=\s*(\d)", text)
    if match:
        ca = int(match.group(1))
        se = int(match.group(2))
        return ca, se
    else:
        return None, None
```

Основная функция обработки сообщений

Это ключевая функция системы, которая:

1. Инициализирует структуру для ответа
2. Обрабатывает контекст (извлекает текст из PDF при необходимости)
3. Определяет намерения пользователя (< CALL> и < SEARCH>)
4. Получает ответ от модели с учетом контекста
5. При необходимости:
 - Строит график (если < CALL> = 1)
 - Ищет документы (если < SEARCH> = 1)
6. Форматирует и выводит ответ

```
def one_mesage(message, context=' '):
    final={'answer':None,'pic':None,'search':None}
    link=extract_local_pdf_links(context)
    if link:
        c=extract_text_from_pdf(link[0])[:3500]
    else:
        c= context
    a = get_intent(message)
    ca,se=parse_call_search(a)
    try:
        answer,df= ask_with_context(message,context=c)
    except:
        answer= ask_with_context(message,context=c)
    final['answer']=answer
    if ca:
        _,pic=call(df,message)
        final['pic']=pic
    if se:
        res=search(message, db)
        final['search']=res[0]

    p1=f"Ответ модели на вопрос: {final['answer']} \n"
    print(p1)

    if final['search']:
        print(final['search'])
        print(db_1[final['search']])
    if final['pic']:
        display(pic)
```


Результат работы функции

работа с текстовой информацией, построение графика и поиск ответа с базы данных

```
[ ] one_mesasge("Найди мне информацию про историю слова аналитик и построй график прибыли и расходов компании", context=test_p)
```

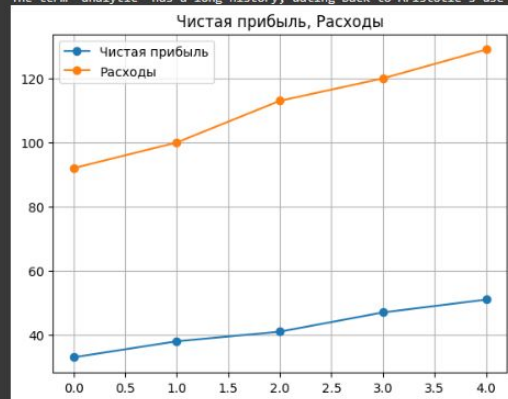
➡ Ответ модели на вопрос: К сожалению, предоставленный текст не содержит информации об истории слова "аналитик". Он описывает финан

По данным текста, динамика прибыли и расходов выглядит следующим образом:

```
* **Выручка:** 125 млн (2019), 138 млн (2020), 154 млн (2021), 167 млн (2022), 180 млн (2023)  
* **Расходы:** 92 млн (2019), 100 млн (2020), 113 млн (2021), 120 млн (2022), 129 млн (2023)  
* **Чистая прибыль:** 33 млн (2019), 38 млн (2020), 41 млн (2021), 47 млн (2022), 51 млн (2023)
```

Document 3

The term "analytic" has a long history, dating back to Aristotle's use in the 4th century BCE to describe logic, and was later re



работа с информацией из PDF-файла

```
[ ] one_mesasge("Проанализируй результаты и дай рекомендации, коротко 5 предложений", context=document_text)
```

➡ Ответ модели на вопрос: Активы ПАО НК «РуссНефть» увеличились с 282 418 млн руб. в 2022 г. до 320 728 млн руб. в 2023 г., что свидетел

Основной цикл диалоговой системы

реализуется интерактивный режим
работы с пользователем

- в цикле запрашивается вопросы и контекст
- для каждого запроса вызывается `one_message()`

завершается при вводе 'q'

```
def main():
    message = input('Введите ваш вопрос (для выхода введите q)')
    context= input('Введите контекст (текст или путь к pdf)')
    while message!='q':
        one_mesasge(message, context=context)
        print('-'*50)
        print('-'*50)
        message = input('Введите ваш вопрос (для выхода введите q)')
        context= input('Введите контекст (текст или путь к pdf)')

main()
```

Результат работы реализованной диалоговой системы

Запрос: “Найди мне информацию про историю слова аналитик и
построй график прибыли и расходов компании”

```
main()
```

Введите ваш вопрос (для выхода введите q) 12 умножить на 20

Введите контекст (текст или путь к pdf)

Ответ модели на вопрос: 12 умножить на 20 равно 240.

Введите ваш вопрос (для выхода введите q) Найди мне информацию про историю слова аналитик и построй график прибыли и расходов компании

Введите контекст (текст или путь к pdf) Финансовый отчёт компании "Альфа Тех" за 2019-2023 гг. показывает устойчивый рост. Выручка ком

Ответ модели на вопрос: К сожалению, в предоставленном тексте нет информации об истории слова "аналитик". Текст содержит только финан

Что касается графика прибыли и расходов, я могу предоставить информацию о них из текста:

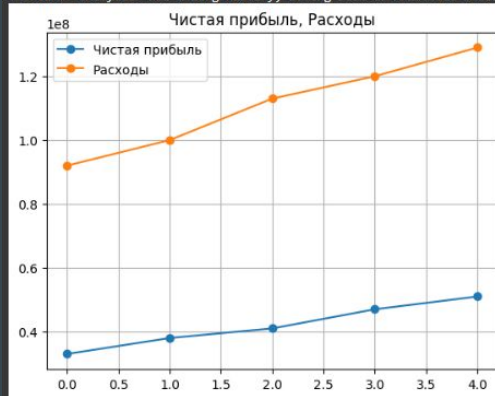
* **Выручка:** 125 млн (2019), 138 млн (2020), 154 млн (2021), 167 млн (2022), 180 млн (2023)

* **Расходы:** 92 млн (2019), 100 млн (2020), 113 млн (2021), 120 млн (2022), 129 млн (2023)

* **Чистая прибыль:** 33 млн (2019), 38 млн (2020), 41 млн (2021), 47 млн (2022), 51 млн (2023)

Document 3

The term "analytic" has a long history, dating back to Aristotle's use in the 4th century BCE to describe logic, and was later referen



Введите ваш вопрос (для выхода введите q) q

Введите контекст (текст или путь к pdf)

Ключевые преимущества системы

1. Интуитивное взаимодействие

- Распознавание запросов на естественном языке
- Простота использования без технической подготовки

2. Автоматизация процессов

- Извлечение данных из документов и отчётов
- Генерация аналитики и визуализаций
- Значительное снижение ручного труда

3. Адаптивность под бизнес-задачи

- Поддержка любых форматов данных
- Быстрая настройка под новые требования

4. Полная конфиденциальность

- Локальное развёртывание
- Защита корпоративных данных

5. Экономия ресурсов

- Ускорение аналитических процессов в 3-5 раз
- Перераспределение персонала на стратегические задачи

Результат: Интеллектуальный инструмент для принятия решений на основе данных с человеко-ориентированным интерфейсом.

Перспективы развития

Система обладает модульной архитектурой для поэтапного масштабирования под задачи бизнеса.

1. Голосовой интерфейс
 - Интеграция распознавания и синтеза речи
2. Работа с корпоративными данными
 - Прямое подключение к SQL-базам
 - Автогенерация запросов из текста
3. Удобные интерфейсы
 - Например Telegram-бот для мобильного доступа

4. Углублённая аналитика

- Расчёт сложных метрик (EBITDA, ROI)
- Прогнозные модели на основе данных

5. Интеллектуальный поиск

- Векторная база знаний
- Расширение базы знаний до десятков тысяч документов

6. Корпоративные функции

- Многопользовательский режим
- Система сессий, ролей с учетом доступа

Сравнительная таблица

Критерий	Gemma 3-12B IT-QAT	Облегчённая модель (2B)
Точность и надёжность	Высокая	Средняя
Работа с PDF и числами	Отлично	Допустимо
Ресурсы	Требует GPU / сервер	Запускается на CPU
Время ответа	Дольше	Быстрее
Обработка сложных запросов	Уверенно	Ограниченно

Дополнительно: пример с облачной моделью gpt-3.5-turbo

```
response = ask_inference_api(
    prompt="ПРОАНАЛИЗИРУЙ ОТЧЕТ СДЕЛАЙ ВЫВОДЫ И ДАЙ СОВЕТЫ, ОТВЕЧАЙ НА РУССКОМ",
    context_messages=[
        {
            "role": "user",
            "content": "'Финансовый отчёт компании \"Альфа Тех\" за 2019-2023 гг. показывает устойчивый рост. Выручка компании увеличи"
        },
    ],
    sys_prompt="Ты — ассистент по бизнес-аналитике."
)

print(response)
```

Анализируя финансовый отчет компании "Альфа Тех", можно сделать следующие выводы:

1. Выручка компании увеличивается ежегодно с 2019 по 2023 год, что говорит о стабильном росте и увеличении объема бизнеса.
2. Рост выручки сопровождается увеличением расходов, но в целом расходы в каждом году покрываются выручкой, приводя к положительной чистой
3. Маржа прибыли по годам также растет, что свидетельствует о повышении эффективности бизнеса.
4. Снижение долговой нагрузки может быть связано с повышением ликвидности компании, поскольку она может увеличивать свой капитал за счет
5. Капитализация компании на конец 2023 года достигла 750 млн рублей, что говорит о положительном развитии бизнеса и увеличении его стабильности.

Выводы: У компании "Альфа Тех" наблюдается стабильный рост выручки и чистой прибыли, повышение эффективности бизнеса, а также снижение долговой нагрузки.

Рекомендации:

1. Для дальнейшего развития компании необходимо продолжать усилия по увеличению выручки и повышению эффективности бизнеса.
2. Необходимо monitoring долговой нагрузки и постоянно улучшать ликвидность компании, чтобы снизить риски возникновения кредитных проблем.
3. Предложить инвесторам информировать о положительном развитии бизнеса и увеличении капитализации, чтобы привлечь дополнительные инвестиции.
4. Для оптимизации бизнес-процессов рекомендуется проводить регулярный анализ и мониторинг финансовой деятельности компании, а также с

Дополнительно: Сравнительная таблица

Критерий	Gemma 3-12B IT-QAT (локально)	Облегчённая модель (2B)	GPT-3.5-Turbo (облако)
Точность и надёжность	Высокая	Средняя	Очень высокая
Работа с PDF и числами	Отлично (при доработке)	Допустимо	Отлично (при наличии контекста)
Ресурсы	Требует мощный GPU / сервер	Запускается на CPU	Требует подключения к API (облачные ресурсы)
Время ответа	Дольше	Быстрее	Среднее (зависит от нагрузки сервиса)
Обработка сложных запросов	Уверенно	Ограниченно	Очень уверенно (лучше всех справляется с многошаговыми и аналитическими задачами)
Стоимость	Бесплатно (при локальном запуске)	Бесплатно	Платно (по количеству токенов)
Простота интеграции	Требует настройки окружения	Простая	Простая (через API)
Конфиденциальность	Высокая (данные не уходят в облако)	Высокая	Ниже (данные отправляются в облако)

Итог

Разработанная интеллектуальная система подтверждает высокую актуальность использования языковых моделей в бизнес-аналитике. Она снижает порог входа в работу с данными, ускоряет принятие решений и демонстрирует потенциал частичной автоматизации аналитической работы. Полученные результаты могут быть основой для дальнейшего развития, включая интеграцию в корпоративные среды и расширение функциональности (например, генерацию отчётов или прогнозных моделей).

Заключение

В рамках данной работы была разработана и реализована автоматическая диалоговая система, способная эффективно взаимодействовать с пользователем на естественном языке в сфере бизнес-аналитики. Система обеспечивает извлечение данных из текстов и PDF-документов, формирует структурированные таблицы, визуализирует показатели и осуществляет семантический поиск по базе аналитических статей.

Для обработки запросов и генерации ответов была выбрана языковая модель **Gemma 3-12B IT-QAT**, которая показала высокую точность, надёжность и способность решать прикладные аналитические задачи локально. Также использовалась облегчённая версия **Gemma 2B** — как альтернатива для менее ресурсоёмкой работы. Для сравнения качества использовалась облачная модель **GPT-3.5-turbo**, что позволило оценить преимущества и ограничения различных архитектур.

Система не только автоматически обрабатывает числовые и текстовые данные, но и делает это в интерактивном режиме с возможностью построения графиков, обработки PDF-файлов и поиска релевантной информации по смыслу запроса.