

## Применение языковых моделей для генерации SQL-запросов в системах оперативной аналитики

Мысив Виталина Владимировна  
бакалавр  
Аркадий Дмитриевич Шнейдер  
ассистент

ФГБОУ ВО «Владивостокский государственный университет»  
Россия. Владивосток  
E-mail: millervitalina2@gmail.com; Тел. +79841471877  
ул. Гоголя, 41, г. Владивосток, Приморский край, Россия, 690014

*Данная работа посвящена исследованию возможностей больших языковых моделей (LLM) для решения задачи преобразования естественного языка в SQL-запросы (NL2SQL) в контексте систем оперативной бизнес-аналитики. В работе анализируется эффективность модели Google Gemini, разрабатывается прототип чат-бота с использованием low-code платформы n8n, и оценивается качество генерации SQL на различных типах запросов. Экспериментально подтверждено, что LLM способны генерировать синтаксически корректные SQL-запросы в 93% случаев, а время ответа составляет 10–20 секунд, что приемлемо для задач оперативной аналитики.*

**Ключевые слова и словосочетания:** NL2SQL, языковые модели, Google Gemini, n8n, low-code, оперативная аналитика, генерация SQL, естественный язык, ИИ-агенты, чат-бот.

### Application of Language Models for SQL Query Generation in Operational Analytics Systems

*This paper investigates the capabilities of large language models (LLMs) for solving the task of natural language to SQL query conversion (NL2SQL) in the context of operational business analytics. The study analyzes the effectiveness of the Google Gemini model, develops a chat bot prototype using the n8n low-code platform, and evaluates SQL generation quality across various query types. The experiment confirms that LLMs can generate syntactically correct SQL queries in 93% of cases, with response times of 10–20 seconds, which is acceptable for operational analytics tasks.*

**Keywords:** NL2SQL, language models, Google Gemini, n8n, low-code, operational analytics, SQL generation, natural language, AI agents, chat bot.

### Введение

Традиционный доступ к данным в компаниях остаётся сложной и трудоёмкой задачей. Бизнес-пользователи - менеджеры, аналитики, руководители - не владеют языком SQL, который необходим для извлечения информации из реляционных баз данных. В результате каждый аналитический запрос требует участия профильного специалиста: менеджер формулирует вопрос на естественном языке, аналитик переводит его в SQL-запрос, выполняет и возвращает результат. По оценкам автора, время получения даже простого отчёта может составлять от нескольких часов до одного рабочего дня. Особенно остро эта

проблема стоит для компаний, где аналитическая нагрузка высока, а скорость принятия решений критически важна.

В связи с этим возникает вопрос: можно ли автоматизировать процесс преобразования естественного языка в SQL-запросы с использованием современных больших языковых моделей (LLM)? Цель настоящего исследования - изучить возможности языковой модели Google Gemini для решения задачи NL2SQL (Natural Language to SQL), разработать экспериментальный прототип интеллектуального помощника и оценить эффективность такого подхода для систем оперативной аналитики.

### **Современное состояние задачи NL2SQL**

NL2SQL - это задача преобразования вопроса, сформулированного на естественном языке, в структурированный SQL-запрос. До недавнего времени для её решения использовались специализированные модели, требовавшие обучения на больших размеченных датасетах, таких как Spider или WikiSQL. Однако с появлением больших языковых моделей (GPT, Gemini, GigaChat) открылись новые возможности: LLM способны генерировать SQL-запросы без предварительного дообучения (zero-shot и few-shot режимы) только на основе текстового описания схемы базы данных и нескольких примеров.

В контексте оперативной бизнес-аналитики NL2SQL-системы должны отвечать следующим требованиям:

- точность - корректное понимание вопроса и генерация семантически верного SQL;
- быстрдействие - время ответа не должно превышать 20–30 секунд;
- поддержка различных типов запросов - агрегация, фильтрация, группировка, топ-N, временная аналитика;
- устойчивость к некорректным запросам - обработка ситуаций, когда пользователь задаёт вопрос не по теме или бессмысленный.

Одним из ключевых факторов успешного применения LLM для NL2SQL является качество промпта - текстовой инструкции, которая описывает модели схему базы данных, правила генерации и примеры правильных запросов. В ходе исследования было установлено, что добавление примеров в промпт повышает точность генерации на 20–30%.

### **Архитектура экспериментального прототипа**

Для проведения эксперимента была разработана архитектура, интегрирующая языковую модель Google Gemini в конвейер оперативной аналитики. Прототип включает следующие компоненты:

- n8n - low-code платформа автоматизации, обеспечивающая оркестрацию всех этапов обработки запроса;
- Google Gemini - большая языковая модель, выполняющая генерацию SQL-запросов;
- PostgreSQL - реляционная база данных, содержащая тестовые данные о продажах;
- Python-сервер - компонент для визуализации результатов в виде графиков.

Пользователь задаёт вопрос на русском языке в чат-боте. Сообщение поступает в n8n, где сначала проходит проверку на осмысленность (фильтрация некорректных запросов). Затем вопрос передаётся в Google Gemini, которая на основе промпта генерирует SQL-запрос. Сгенерированный запрос выполняется в PostgreSQL, а результат возвращается пользователю в виде текста и, при необходимости, графика. На рисунке 1 представлен пример визуального workflow в n8n.

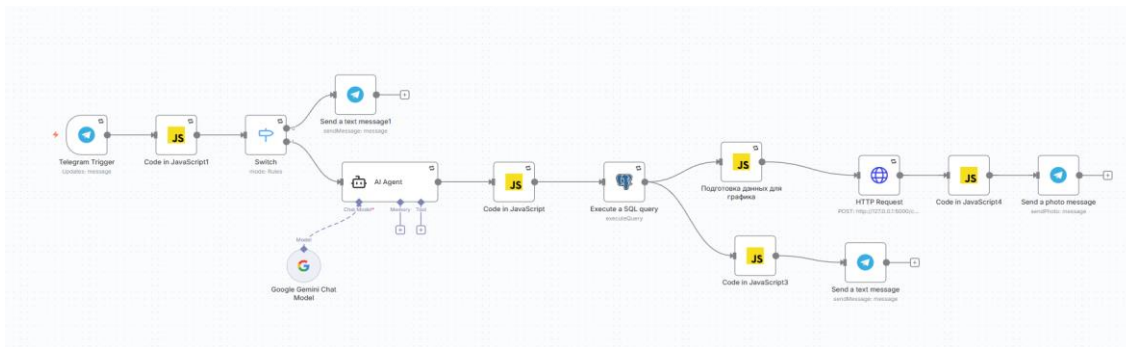


Рис. 1. Визуальный рабочий процесс ИИ-агента в n8n для генерации SQL-запросов

### Методика эксперимента

Для оценки качества работы прототипа была составлена тестовая выборка из 15 вопросов, покрывающих различные типы аналитических запросов. Распределение вопросов по типам и сложности представлено в таблице 1.

Таблица 1

### Типы тестовых запросов для оценки качества генерации SQL

№	Тип запроса	Пример вопроса	Ожидаемая сложность
1	Агрегация	«Покажи общую сумму продаж»	Низкая
2	Группировка	«Продажи по регионам»	Низкая
3	Топ-N	«Топ-3 региона по сумме продаж»	Низкая
4	Фильтрация по одному полю	«Продажи в Москве»	Средняя

Для каждого типа запроса было составлено по 2–3 вопроса, что обеспечило репрезентативность выборки. Всего было сформулировано 15 тестовых вопросов, охватывающих основные сценарии оперативной аналитики: от простых агрегирующих запросов до сложных комбинированных условий с фильтрацией по нескольким полям и временными ограничениями.

Оценка производилась по трём критериям:

- синтаксическая корректность - насколько правильно сгенерированный SQL с точки зрения синтаксиса PostgreSQL;
- семантическая точность - насколько SQL соответствует смыслу вопроса;
- время выполнения - время от отправки вопроса до получения ответа (включая генерацию SQL, выполнение запроса и формирование ответа).

### Результаты эксперимента

В ходе эксперимента были получены следующие результаты.

Синтаксическая корректность составила 93%: из 15 запросов только один содержал синтаксическую ошибку. Ошибка возникла при обработке сложного комбинированного запроса с несколькими условиями фильтрации.

Семантическая точность составила 87%: в большинстве случаев сгенерированный SQL правильно отражал смысл вопроса. Основные ошибки были связаны со сложными комбинированными запросами и распознаванием русскоязычных синонимов (например, «выручка» вместо «сумма продаж»).

Среднее время генерации SQL составило от 3 до 5 секунд в зависимости от сложности запроса. Полное время ответа (включая выполнение запроса в PostgreSQL и формирование ответа) - 10–20 секунд.

В таблице 2 представлена детализация результатов по типам запросов.

Таблица 2

**Таблица 2. Результаты работы ИИ-агента на различных типах запросов**

№	Тип запроса	Синтаксическая корректность	Семантическая точность	Время генерации (сек)
1	Агрегация	100%	100%	2,8
2	Группировка	100%	100%	3,0
3	Топ-N	100%	100%	3,1
4	Фильтрация по одному полю	100%	90%	3,5
5	Фильтрация по двум полям	100%	90%	3,8
6	Временная аналитика	100%	85%	4,0
7	Комбинированный запрос	80%	80%	4,5

Также была реализована функция обработки некорректных запросов. Если пользователь вводит бессмысленный запрос («Сколько?», «Привет», «асфжывароолдж»), бот не передаёт его в языковую модель, а возвращает сообщение с просьбой переформулировать вопрос и примерами корректных запросов. Это повышает удобство использования и предотвращает генерацию ошибочных SQL-запросов.

## Выводы

Проведённое исследование позволяет сделать следующие выводы.

Первый вывод: современные большие языковые модели, в частности Google Gemini, способны эффективно решать задачу NL2SQL без дополнительного обучения. 93% синтаксически корректных запросов подтверждают, что модель хорошо справляется с типовыми аналитическими запросами.

Второй вывод: качество генерации критически зависит от качества промпта. Добавление в промпт описания схемы базы данных и 5–7 примеров правильных запросов повышает точность генерации на 20–30% по сравнению с простым промптом без примеров.

Третий вывод: время ответа (10–20 секунд) является приемлемым для систем оперативной аналитики, где приоритетом является не сверхсекундная скорость, а возможность получить данные без участия аналитика.

Четвёртый вывод: разработанный прототип подтверждает возможность создания интеллектуальных помощников для работы с базами данных. Внедрение такого решения в компании позволит менеджерам получать аналитику без знания SQL, а аналитикам - сосредоточиться на более сложных задачах.

## Перспективы применения

Разработанный прототип и предложенный подход открывают широкие возможности для практического применения:

- бизнес-аналитика - создание корпоративных чат-ботов для оперативного получения отчётов по продажам, клиентам, финансам;
- студенческие проекты и образование - использование платформы n8n и LLM для быстрой разработки учебных проектов без глубокого знания программирования;
- автоматизация запросов к базам данных - интеграция с корпоративными мессенджерами (Slack, Teams) для обеспечения доступа к данным широкого круга сотрудников;
- интеграция с российскими LLM - замена Google Gemini на GigaChat или YandexGPT для использования в организациях с требованиями к импортозамещению.

В перспективе предполагается расширение исследований в следующих направлениях:

- сравнительный анализ различных LLM (Gemini, GigaChat, GPT) для русскоязычных NL2SQL-запросов;
- использование fine-tuning для повышения точности генерации в узких предметных областях;
- добавление поддержки более сложных типов запросов (многотабличные JOIN, подзапросы, оконные функции).

## Заключение

Технологии больших языковых моделей продолжают активно развиваться, и задача преобразования естественного языка в SQL-запросы становится всё более доступной для практического применения. В ходе исследования был разработан прототип чат-бота, интегрирующего Google Gemini в конвейер оперативной аналитики. Экспериментально подтверждено, что LLM способны генерировать синтаксически корректные SQL-запросы в 93% случаев, а время ответа составляет 10–20 секунд.

Разработанный прототип может быть использован в качестве основы для создания корпоративных систем оперативной аналитики, а предложенная архитектура - тиражирована для других предметных областей. Дальнейшие исследования будут направлены на повышение точности генерации сложных запросов и интеграцию с российскими языковыми моделями и сервисами.

- 
1. Google AI Studio. Gemini API Documentation [Электронный ресурс] - URL: <https://ai.google.dev/gemini-api/docs> (дата обращения: 10.04.2026).
  2. Документация платформы [Электронный ресурс] - URL: <https://docs.n8n.io>
  3. Liu B., et al. A Survey of Natural Language to SQL (NL2SQL) Systems // arXiv preprint arXiv:2310.09187, 2023.