

## ПАО «Газпром Нефть»

**Разработка системы для ответа на запросы клиентов, в которой ответы формируются с помощью подгруженных заранее документов**

## Описание задачи

Вам необходимо разработать систему для ответа на запросы клиентов, в которой ответы формируются с помощью подгруженных заранее документов. Система должна уметь извлекать вопросы из сложных и объемных запросов, находить нужный контекст для ответа и генерировать финальный ответ.

Сотрудники методологической функции каждый день обрабатывают запросы по разъяснению методологии и процедур компании, и дают письменные ответы на запросы. Они формируют ответы на свободные запросы пользователей, основываясь на 5 заданных документах (нормативные и юридические документы по закупкам, федеральные законы, приказы компании).

## Проблема

Много времени сотрудников компании уходит на обработку запросов: зафиксировать контекст письма, понять вопросы в письме, определить, где искать ответ на вопрос.

Запросы требуют сложного анализа и непростых ответов.

## Основные этапы решения

Разработать систему RAG, которая:

1. Обрабатывает текстовый запрос/письмо содержащее вопросы, с большим контекстом или вступительной частью;
2. Извлекает из текста набор релевантных вопросов, которые должны отражать цель и суть письма, но должны быть пригодны для передачи в ретривер;
3. Находит релевантную информацию в векторной базе знаний;
4. Генерирует ответ на основе запроса (в т.ч. вопросов из запроса).

## Необходимая и доступная для решения информация

Вам будут доступны

1. набор из 3 документа: 2 word и 1 pdf;  
<https://drive.google.com/drive/folders/19P7PnZrp5geCdeTMrGf4YriRC59DOUet?usp=sharing>
2. запросы/письма, которые поступали сотрудникам компании по 5 документам. Запросы/письма содержат большой контекст или вступительную часть.  
[https://docs.google.com/spreadsheets/d/1Gc5N8DiMcj1gmisO9lj8L\\_8HaiKPBHVR/edit?usp=sharing&oid=117351453470675749663&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1Gc5N8DiMcj1gmisO9lj8L_8HaiKPBHVR/edit?usp=sharing&oid=117351453470675749663&rtpof=true&sd=true)

## Желаемый результат

Telegram-бот, принимающий на вход текстовые сообщения, отвечает основываясь только на загруженных документах.

Ответ от бота содержит:

- Извлеченные вопросы
- Подходящие чанки текста из базы знаний
- Финальный ответ

## Перспективы развития проекта

Результаты могут стать частью решения, которая разрабатывается внутренней командой. ИИ- помощник для поддержки внутренних сотрудников и поставщиков для ответа на запросы по процедурам закупки

## Дополнительная информация

### Стек технологий (требования)

- Наличие репозитория с воспроизводимым кодом
- Python версии 3.8+ и любые библиотеки с открытым исходным кодом
- Языковые модели: yandexgpt, gigachat и любые открытые LLM и OCR
- Docker
- Работающий телеграм-бот
- API для ответа на запросы

## Прочие условия: NDA, передача результата, взаимодействие с заказчиком

Будет поддержка внутренней команды Центра экспериментальных решений Газпромнефть-Снабжения:

- Гаджаева Лаура [Gadzhaeva.LR@gazprom-neft.ru](mailto:Gadzhaeva.LR@gazprom-neft.ru) (Руководитель центра)
- Илья Молчанов [Molchanov.IA@gazprom-neft.ru](mailto:Molchanov.IA@gazprom-neft.ru) (Руководитель проекта )  
**ответственный за коммуникацию по задаче**
- Владимир Панов [Panov.VO@gazprom-neft.ru](mailto:Panov.VO@gazprom-neft.ru) (Дата-сайентист)

## Критерии оценки решения

### 1. Соответствие решения поставленной задаче

Решение отвечает на вопросы, используя предоставленные документы, реализовано в виде телеграм бот. Максимальный балл 5.

### 2. Качество модели

Для тестирования используется cross encoder, оценка происходит по следующей формуле:

$$F = \frac{1}{N} \sum_i R_{crossencoder}(y_{true,i}, y_{pred,i}),$$

где  $y_{true}$  - это эталонные ответы,  $y_{pred}$  - это ответы вашего решения.

Пример расчета скоры для одной пары

<https://colab.research.google.com/drive/1t6enEtQC0iE9CD3JPi0zbPbzpw4tfOX7?usp=sharing>

Проверка решения пройдет через два тестирования (участники хакатона отправляют свои результаты организаторам):

#### Тестирование во время хакатона

- На время хакатона дается 3 попытки для теста
- Открытый тест с разной сложностью запросов, который не меняется во время хакатона и открыт для всех участников

#### Контрольное/итоговое тестирование:

- Дается 1 попытка
- Закрытый тест, участники хакатона не знакомы с запросами

Оценка вычисляется по среднему между максимальным баллом из 3 попыток и итоговым баллом. Максимальный балл за каждую попытку 15 баллов, суммарно максимум 15 баллов.

### 3. Качество презентации

- Указаны цель проекта, задачи, решаемые продуктом, указаны критерии качества продукта, результаты описаны детально и структурировано

Максимальный балл 5 баллов.

#### **4. Оценка технического решения**

- Система должна разворачиваться внутри Docker контейнера или контейнеров
- README с инструкцией по разворачиванию каждого модуля системы
- Желательно использование аннотирование типов
- Код передаётся либо в виде GIT репозитория или zip архива
- Файл с зависимостями

Максимальный балл 10 баллов

#### **5. UX/UI телеграм бота**

Максимальный балл 5 баллов.

Итоговый балл по всем критериям будет разделен на 2.

*Желаем успехов!*

## Примеры запросов

### Письмо 1:

От инициатора поступил вопрос, просим проконсультировать по следующему вопросу: по ранее утвержденной закупке услуг, договор был направлен на согласование в Тесса, в ходе согласования поступили замечания от ДЭИКП по п.5.4

- 1) если по договору предполагается работа по гарантированному/негарантированному объему, то категорию дд необходимо указать "стоимостной".
- 2) В приложение 1 добавить формулировку о том, что негарантированный объем переводится в гарантированный в объеме (сумма) на (2023/2024/2025) год.

Так как договор является рамочным, но формулировка п.5.4. не подходит под категорию рамочного, специалист ДЭИКП согласовал замену п.5.4. из договора:

5.4. Состав услуг по договору состоит из негарантированного объема. Выполнение негарантированного объема услуг осуществляется на основании потребности Заказчика, при наличии источника финансирования и оформляется путем заключения Приложения (форма Приложения №1 к настоящему Договору) к настоящему Договору, подтверждающего перевод негарантированного объема услуг в гарантированный. До момента подписания сторонами такого Приложения в отношении негарантированного объема или его части у Исполнителя не возникнет обязанность оказать услуги в такой части негарантированного объема, а у Заказчика не возникнет обязанность по их оплате.

На формулировку соответствующую рамочному договору:

5.4. Состав, требования, стоимость, функциональный объем, формы и сроки предоставления результатов услуг фиксируются в приложениях (форма Приложения №1) к настоящему договору. Заказчик обязуется оплатить Исполнителю стоимость Услуг в суммах, указанных в соответствующих Приложениях к настоящему Договору. Выполнение объема услуг осуществляется на основании потребности Заказчика, при наличии источника финансирования.

Подскажите, на сколько это критично для договора прошедшего закупочную процедуру и согласование закупки?

### **Письмо 2:**

Коллеги, добрый день!

Прошу проконсультировать по вопросам проведения закупок с целью заключения рамочных договоров на СМР и текущий ремонт:

Вопросы следующие:

1. По какой методике ценообразования можно заключить рамочный договор на выполнение работ по обустройству/модернизации /дооснащению офисных помещений в условиях отсутствия РД и сметной документации на работы?

Исходные данные такие:

- Ориентировочная площадь офисных помещений суммарно – до 10 000 м<sup>2</sup> в год,
- состав работ – перепланировка помещений, отделочные работы, модернизация внутренних инженерных систем
- сроки выполнения работ по каждому объекту – от 2 до 5 месяцев (ориентировочно),
- офисные помещения расположены в СПб, Москве, Сочи и возможно в других регионах, где есть офисы ГПН.

2. Планируется заключение рамочного договора на текущий ремонт– мелкие работы по ремонту офисных помещений, ремонту инженерных сетей в действующих офисах.

Подготовлен прайс с расценками на предполагаемые виды работ, потребность в которых была выявлена при эксплуатации объектов. При этом есть риск, что прайс будет неполным, и возникнет потребность в выполнении работ, которых нет в прайсе.

Как можно определить цену таких работ:

- зафиксировать в договоре часовую тарифную ставку, например
- зафиксировать условие, что стоимость будет определена по действующей нормативной базе (ТЕР, ФЕР с
- применением методики Минстроя)