**Задание на llm.**

Сделать систему DOCQA на базе КОАП города Москвы.

Требования (рекомендация):

Фреймворк - Langchain;

Интерфейс – Gradio;

Память – Окно;

Модель – Llama(Сайга) GGML;

Разработать и использовать Агента.

Ожидаемый результат:

Ответ бота должен содержать ответ на вопрос, номер статьи, референсный фрагмент.