

Определение поисковых критериев для глубоко залегающих горизонтов нефти и газа Волго-Уральского нефтегазоносного бассейна с помощью больших языковых моделей.

Научный руководитель – Карпушин Михаил Юрьевич

Болдырев Арсений Александрович

Студент (магистр)

Московский государственный университет имени М.В.Ломоносова, Геологический факультет, Кафедра геологии и геохимии горючих ископаемых, Москва, Россия

E-mail: boldyrev.ars.al.ie@gmail.com

За последние десятилетия эффективность поисков нефти и газа в традиционных интервалах глубин существенно снизилась. В зрелых нефтегазоносных провинциях основная часть крупных залежей уже открыта, что смещает фокус геологоразведочных работ в сторону глубокозалегающих горизонтов. Однако их изучение сопряжено с высокой неопределённостью и значительными затратами, а анализ накопленной геологической информации требует обработки больших объёмов разрозненных текстовых и фондовых материалов.

Целью работы является оценка возможности применения больших языковых моделей (LLM) в составе RAG-архитектуры (Retrieval-Augmented Generation) для систематизации геологических данных и выявления критериев перспективности глубоких горизонтов.

Входными данными системы являются текстовые документы различного формата (PDF, DOC, сканированные отчёты после OCR). На этапе предобработки выполняются очистка, разбиение на фрагменты (чанки, пассажи) фиксированного или динамического объёма и извлечение метаданных (глубина, стратиграфия, регион). Каждый фрагмент преобразуется в векторное представление с использованием модели эмбедингов и сохраняется во векторной базе данных. При поступлении запроса он также преобразуется в вектор и сопоставляется с индексами из базы. Модуль поиска возвращает N наиболее семантически близких фрагментов. Эти фрагменты формируют контекст, который подаётся на вход LLM вместе с исходным вопросом.

Таким образом, модель генерирует ответ, основываясь не только на заложенных в нее в процессе обучения данных, а формирует ответ на основе найденных дополнительных данных. Выходом системы является структурированный текст, содержащий обобщённые критерии, ссылки на источники и интерпретацию выявленных признаков. Дополнительно возможна генерация перечня факторов перспективности в виде списка или таблицы.

Апробация выполнена на материалах по глубоким объектам Волго-Уральской провинции и Карачаганакскому месторождению. Система устойчиво выявляет повторяющиеся признаки перспективности.

Предложенная RAG-архитектура позволяет перейти от фрагментарного анализа документов к формированию целостной концептуальной модели глубокой залежи. Использование LLM в качестве интерпретационного ядра при сохранении контроля через модуль поиска снижает риск галлюцинаций и повышает обоснованность выводов. Подход представляет перспективное направление цифровизации нефтегазовой отрасли и может использоваться как инструмент поддержки принятия решений при оценке глубоких горизонтов.