Секция «2.2 Биоинформатика»

Использование больших языковых моделей и архитектуры RAG с количественными признаками изображений для интерпретации цитологических анализов щитовидной железы

Научный руководитель – Зайцев Константин Сергеевич

Основин С.С.[1], *Садохин А.А.*[2]

1 - Национальный исследовательский ядерный университет «МИФИ», Москва, Россия, *E-mail: 1300stas1300@gmail.com*; 2 - Национальный исследовательский ядерный университет «МИФИ», Факультет кибернетики и информационной безопасности, Москва, Россия, *E-mail: alexeisadohin@yandex.ru*

Целью данного исследования выступает создание инструмента для медицинского работника, позволяющий формировать клиническую трактовку, интерпретацию либо аргументацию выводов внешних систем, основываясь на объединении числовых параметров анализов и упорядоченных профессиональных знаний. Для этого мы используем возможности анализа контекста больших языковых моделей и механизмы RAG. Это позволяет формировать ответы опирающиеся на достоверные факты [4].

Реализованный нами механизм RAG состоит из двух частей. Первая часть представляет собой подготовленную и предобработанную базу знаний, составленную на базе официальных документов связанных с TBSRTC [1]. Вторая часть – числовые характеристики, полученные внешней моделью сегментации из цитологических изображений (количество клеток и типы клеточных образований, пропорция и т. д.).

Архитектура работы системы включает следующие компоненты: модуль семантического поиска (эмбеддинговая модель с индексацией в FAISS) для извлечения релевантных фрагментов из корпуса TBSRTC, модуль формирования контекстного промпта, генеративное ядро на базе открытых LLM (тестирование проведено на "Qwen2.5-7B-Instruct", "Llama-3.1-8B-Instruct", "Gemini 2.5 Pro", "GigaChat-Pro", "DeepSeek-LLM-7B-Chat") [3].

Проверка и валидация системы выполнена на выборке обезличенных клинических случаев, предоставленных НМИЦ эндокринологии им. академика И.И. Дедова Минздрава России. Помимо использования стандартных метрик для определения количества правильных постановок меток Bethesda (F1-Score, Recall и Accuracy), для оценки работы системы RAG и генеративных моделей использовались метрики ROUGE, Perplexity, Context Relevance, Faithfulness [2].

Главным достижением является анализ и сравнение работы различных LLM моделей с подаваемым на вход контекстом, а также снижение галлюцинаций данных моделей. Ключевым выводом является тот факт, что без использования модуля для извлечения контекста, доля некорректных ответов моделей с количеством параметров менее 7B резко возрастает. В дальнейшей перспективе планируется расширить перечень анализируемых методов RAG в подборе текстовых фрагментов, а также масштабировать предложенное решение на другие медицинские домены.

Источники и литература

1) Ali S.Z., Baloch Z.W., Cochand-Priollet B. et al. The 2023 Bethesda System for Reporting Thyroid Cytopathology // Thyroid. 2023. Vol. 33, No. 9. P. 1039–1044. doi: 10.1089/thy.2023.0141

2) Gu B., Desai R.J., Lin K.J. et al. Probabilistic medical predictions of large language models // npj Digital Medicine. 2024. Vol. 7, No. 367. doi: 10.1038/s41746-024-01366-4

3) Gupta S. A comprehensive survey of retrieval-augmented generation (RAG): evolution, current landscape and future directions / S. Gupta, R. Ranjan, S.N. Singh // arXiv preprint. — 2024

4) Izacard G. Atlas : few-shot learning with retrieval augmented language models / G. Izacard, P. Lewis, M. Lomeli et al. // arXiv preprint. — 2022